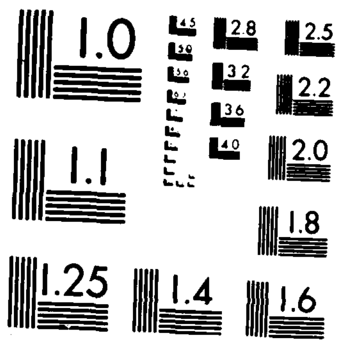
AD-A174 825 EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES IF THEY ARE ACTUA (U) OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS M L MOESCHBERGER ET AL 26 NOV 85 1/2

UNCLASSIFIED AFOSR-TR-86-2042 AFOSR-82-0307 F/G 14/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A174 825

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR. 86-2042
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Effects of Assuming Independent Component Failure Times, If They Are Actually Dependent, in a Series System
5. TYPE OF REPORT & PERIOD COVERED: Annual Report 10/1/84 to 10/31/85
6. PERFORMING ORG. REPORT NUMBER: 714837/763265
7. AUTHOR(s): Melvin L. Moeschberger and John P. Klein
8. CONTRACT OR GRANT NUMBER(s): Grant No. AFOSR 82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The Ohio State University Research Foundation, 1314 Kinnear Road, Columbus, Ohio 43212
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: ~~92~~ 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: AFOSR-NM, Building 410, Bolling AFB, Washington, D.C. 20332
12. REPORT DATE: November 26, 1985
13. NUMBER OF PAGES:
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): AFOSR-PKZ, Building 410, Bolling AFB, Washington, D.C. 20332 nm
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited

DTIC ELECTE DEC 5 1986 A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): modeling series systems, system reliability, competing risks, bivariate exponential distributions, test for independence, consistency of the product limit estimator, dependent risks, estimation of component life
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

The overall objective of this proposal is to investigate the robustness to departures from independence of methods currently in use in reliability studies when competing failure modes or competing causes of failure associated with a single mode are present in a series system. The first specific aim is to examine the error one makes in modeling a series system by a model which assumes statistically independent component lifetimes when in fact the component lifetimes follow some multivariate distribution. The second specific aim is to

(continued)

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

267360 SH

AFOSR-TR- 86-2042

Annual Report for the
Air Force Office of Scientific Research

by

The Ohio State University Research Foundation
1314 Kinnear Road
Columbus, Ohio 43210

for

Melvin L. Moeschberger
Associate Professor of Preventive Medicine

and

John P. Klein
Assistant Professor of Statistics

Title: Effects of Assuming Independent Component Failure Times, if They are Actually Dependent in a Series System

Period Covered: October 1, 1984 to October 31, 1985

Principal Investigators: Melvin L. Moeschberger and John P. Klein

Phone: (614) 421-3878 or (614) 422-4017

Business Contact: Phillip V. Spina

Phone: (614) 422- 3730

Date of Submission: October 26, 1985

Melvin L. Moeschberger
Melvin L. Moeschberger, Ph.D.

Martin D. Keller, M.D., Ph.D.
Chairperson of Preventive Medicine

John P. Klein, Ph.D.

Jagdish Rustagi, Ph.D.
Chairperson of Statistics

86 12 04 042

Moeschberger, Melvin L.

A. Table of Contents

Topic	Pages
Technical Section	
Abstract	3
Specific Objectives	4
Introduction to Problem and Significance of Study	5-7
Progress Report on Third Year's Work	8
Methods	9
Literature Cited	10-12
Appendices A,B,C,D,E	13

A-1

B. TECHNICAL SECTION

I. Abstract

The overall objective of this proposal is to investigate the robustness to departures from independence of methods currently in use in reliability studies when competing failure modes or competing causes of failure associated with a single mode are present in a series system. The first specific aim is to examine the error one makes in modeling a series system by a model which assumes statistically independent component lifetimes when in fact the component lifetimes follow some multivariate distribution. The second specific aim is to assess the effects of the independence assumption on the error in estimating component parameters from life tests on series systems. In both cases, estimates of such errors will be determined via mathematical analysis and computer simulations for several prominent multivariate distributions. A graphical display of the errors for representative distributions will be made available to researchers who wish to assess the possible erroneous assumption of independent competing risks. A third aim is to tighten the bounds on estimates of component reliability when the risks belong to a general dependence class of distributions (for example, positive quadrant dependence, positive regression dependence, etc.). Major decisions involving reliability studies, based on competing risk methodology, have been made in the past and will continue to be made in the future. This study will provide the user of such techniques with a clearer understanding of the robustness of the analyses to departures from independent risks, an assumption commonly made by the methods currently in use.

II. Specific Objectives

The overall objective is to investigate the robustness to departures from independence to methods currently in use in reliability studies when competing failure modes or competing causes of failure associated with a single mode are present in a series system. We shall also refer to such competitive events as competing risks. The approach will be through the investigation of certain aspects of specific parametric multivariate distributions or by classes of distributions which are appropriate in reliability analyses when there are competing risks present.

The specific objectives are:

1) to assess the error incurred in modeling system life in a series system assumed to have independent component lifetimes when in fact the component lifetimes are dependent.

2) to assess the error in estimating component parameters (i.e., component reliability, mean component life, etc.) in a series system employing either parametric or nonparametric models which assume independent component failure times when in fact the lifetimes are dependent and follow some plausible multivariate distribution.*

3) to derive bounds on component reliability when the failure modes are dependent and fall in a particular dependence class (e.g., positive quadrant dependence, positive regression dependence, etc.).

4) to develop tests of independence, based on data collected from series systems, by making some restrictive assumption about the structure of the systems.**

* A plausible parametric multivariate distribution will be one that satisfies one of the following conditions:

i) the distribution of the minimum of the component failure times closely approximates widely accepted families of system life distributions.

or ii) the marginal distributions closely approximate the distributions of component failure times in the absence of other failure modes.

**This objective has been added to the original objectives because it answers a natural question raised by our preliminary investigation.

III. Introduction to Problem and Significance of Study

Alvin Weinberg (1978) in an editorial comment in the published proceedings of a workshop on Environmental Biological Hazards and Competing Risks noted that "the question of competing risks will not quietly go away: corrections for competing risks should be applied routinely to data." The problem of competing risks commonly arises in a wide range of experimental situations. Although we shall confine our attention in the following discussion to those situations involving series systems in which competing failure modes or competing causes of failure associated with a single mode are present, it is certainly true that we might just as easily speak of clinical trials, animal experiments, or other medical and biological studies where competing events interrupt our study of the main event of interest (cf. Lagokos (1979)).

Consider electronic or mechanical systems, such as satellite transmission equipment, computers, aircraft, missiles and other weaponry consisting of several components in series. Usually each component will have a random life length and the life of the entire system will end with the failure of the shortest lived component. We will examine two situations more closely in which competing risks play a vital role.

First, suppose we are attempting to evaluate system life from knowledge of the individual component lifetimes. Such an evaluation will utilize either an analysis involving mathematical statistics or a computer simulation. At a recent conference on Modeling and Simulation, McLean (1981) presented a scheme to simulate the life of a missile which consisted of many major components in series. The failure distribution associated with each component was assumed to be known (usually exponential or Weibull.) To arrive at the system failure distribution, the components were assumed to act independently of each other. Realistically, this may or may not be the case. If the component lifetimes were dependent for any reason, the computed system failure distribution (as well as its subsequent parameters such as system mean life and system reliability for a specified time) would only crudely approximate the true distribution. The first specific aim of this proposal is to ascertain the error incurred in modeling system life in a series system assumed to have independent component lifetimes (i.e., risks) when, in fact, the risks are dependent.

Second, suppose we wish to evaluate some aspect of the distribution of a particular failure mode based on a typical life test of a series system. The response of interest is the time until failure of a particular mode of interest. Frequently this response will not be observable due to the occurrence of some other event which precludes failure associated with the mode of interest. We shall term such competing events which interrupt our study of the main failure modes of interest as competing risks.

Competing risks arise in such reliability studies when

1) the study is terminated due to a lack of funds or the pre-determined period of observation has expired (Type I censoring).

2) the study is terminated due to a pre-determined number of failures of the particular failure mode of interest being observed (Type II censoring).

3) some systems fail because components other than the one of interest malfunction.

4) the component of interest fails from some cause other than the one of interest.

In all four situations, one may think of the main event of interest as being censored, i.e., not fully observable. In the first two situations, the time to occurrence of the event of interest should be independent of the censoring mechanism. In such instances, the methodology for estimating relevant reliability probabilities has received considerable attention (cf. David and Moeschberger (1978), Kalbfeish and Prentice (1980), Elandt-Johnson and Johnson (1980), Mann, Schafer, Singpurwalla (1974) and Barlow and Proschan (1975) for references and discussion). In the third situation, the time to failure of the component of interest may or may not be independent of the failure times of other components in the system. For example, there may be common environmental factors such as extreme temperature which may affect the lifetime of several components. Thus the question of dependent competing risks is raised. A similar observation may be made with respect to the fourth situation, viz., failure times associated with different failure modes of a single component may be dependent. For a very special type of dependence, the models discussed by Marshall-Olkin (1967), Langberg, Proschan and Quinzy (1978), and Langberg, Proschan, and Quinzy (1981) allow one to convert dependent models into independent ones.

If no assumptions whatever are made about the type of dependence between the distribution of potential failure times, there appears to be little hope of estimating relevant component parameters. In some situations, one may be appreciably misled (cf. Tsiatis (1975), Peterson (1976)). However, as Easterling (1980) so clearly points out in his review of Birnbaum's (1979) monograph

> "there seems to be a need for some robustness studies. How far might one be off, quantitatively, if his analysis is based on incorrect assumptions?"

The second specific aim will address this important issue. First if a specific parametric model which assumes independent risks has been used in the analysis, it would be of interest to know how the error in estimation is affected by this assumption of independence. That is, if independent specific parametric distributions are assumed for the failure times associated with different failure modes when we really should use a bivariate (or multivariate) distribution, then what is the magnitude of the error in estimating component parameters? Secondly, one may wish to allow for a less stringent type of model assumption, and ask the same question with regard to the estimation error. That is, if a nonparametric analysis is performed, assuming independent risks, when some types of dependencies may be present, then what is the magnitude of the estimation error?

The third specific aim will attempt to obtain bounds on the component reliability when the failure times belong to a broad dependence class (e.g., association, positive quadrant dependence, positive regression dependence, etc.). More details will be presented in the methods section.

In summary, competing risk analyses have been performed in the past and will continue to be performed in the future. This study will provide the user of such techniques with a clearer understanding of the robustness to departures from independent risks, an assumption which most of the methods currently in use assume.

IV. Progress Report on Third Year's Work

A summary of the first and second year's work is reported in the annual reports dated October 26th 1983 and October 26th 1984 respectively. We believe that during the past three years we have made substantial progress in dealing with the objectives as outlined on page 4. In addition to the papers and articles referred to in the first two annual reports, we would like to mention some of the more recent work.

First, the recently published paper which investigates the problem of improving the product-limit estimator of Kaplan and Meier (1958) when there is extreme independent right censoring is presented in Appendix A. This paper looks at several techniques for completing the product limit estimator by estimating the tail probability of the survival curve beyond the largest observed death time. Two methods are found to work well for a variety of underlying distributions. The first method replaces those censored observations larger than the biggest death time by the expected order statistics, conditional on the largest death, computed from a Weibull distribution. The Weibull is chosen since it is known to be a reasonable model for survival in many situations. Parameters of the model are estimated in several ways, but the method of maximum likelihood seems to provide the best results. The second method replaces the constant value of the product limit estimator beyond the last death time by the tail of a Weibull survival function. Again parameters are estimated by a variety of methods with the maximum likelihood estimators performing the best.

Second, a paper which obtains bounds on the component reliability, based on data from a series system, for the Oakes (1982) model has been revised. Since this model has the same dependence structure as the random effects model with w having a gamma distribution, these bounds are good for a general class of distributions. The bounds, which are determined by specification of a range of coefficients of concordance, are found by solving a differential equation in the observable system reliability and crude life on one hand and the unobservable component survival function on the other hand. This revision is reproduced in Appendix B.

Third, we have submitted an overview paper for publication which summarizes some of the work performed during the past three years. The results of this paper were presented in an invited talk to the Eastern North American Region of the Biometrics Society at Raleigh, North Carolina in the Spring of 1985. (See Appendix C for a copy of this paper.)

Finally, a paper has been developed which discusses some general properties of a random environmental stress model. Estimation of parameters under the Gamma stress model is considered, and a new estimator based on the scaled total time on test transform is presented (See Appendix D). Part of the results in this paper were presented at the International Statistical Institute meeting in Amsterdam in August, 1985. A copy of that contributed paper is found in Appendix E.

V. Methods

We refer to pages 8-52 of the original proposal for a discussion of the general methodology.

LITERATURE CITED

Ahmed, A.N., Langberg, N.A., Leon, R.V. and Proschan, F. (1978). Two concepts of positive dependence, with applications in multivariate analysis. Statistics Report No. M486, Florida State University, Tallahassee.

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing - Probability Models. Holt, Rinehart and Winston, Inc., New York.

Basu, A.P. and Klein, J.P. (1982). Some recent results in competing risks. Survival Analysis. Crowley and Johnson, Editors, 216-229.

Berman, S.N. (1963). Notes on extreme values, competing risks, and semi-Markov processes. Ann. Math. Statist. 34, 1104-1106.

Birnbaum, Z.W. (1979). On the Mathematics of Competing Risks. U.S. Dept. of HEW. Publication No. 79-1351.

Chiang, C.L. (1968). Introduction to Stochastic Processes in Biostatistics. Wiley, New York.

Cox, D.R. (1962). Renewal Theory. Methuen, London.

David, H.A. (1974). Parametric approaches to the theory of competing risks. Pp. 275-290 in: Proschan and Serfling (1974).

David, H.A. and Moeschberger, M.L. (1978). Theory of Competing Risks. Griffin, London.

Easterling, R.G. (1980). Book review of Z.W. Birnbaum monograph. Technometrics 22: 131-132.

Elandt-Johnson, R.C. and Johnson, N.L. (1980). Survival Models and Data Analysis. Wiley, New York.

Fisher, L. and Kanarek, P. (1974). Presenting censored survival data when censoring and survival times may not be independent. Reliability and Biometry, Statistical Analysis of Lifelength. Siam, 303-326.

Fishman, G.S. (1973). Concepts and Methods in Discrete Event Digital Simulation. Wiley, New York.

Friday, D.S. and Patel, G.P. (1977). A bivariate exponential model with applications to reliability and computer generation of random variates. Theory and Applications of Reliability, Vol. I. C.P. Tsokos and I.N. Shimi, Editors, 527-548.

Gail, M. (1975). A review and critique of some models used in competing risk analysis. Biometrics 31, 209-222.

Gibbons, J.D. (1971). Nonparametric Statistical Inference. McGraw-Hill Book Company, New York.

...

Gumbel, E.J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55: 698-707.

Hoel, D.G. (1972). A representation of mortality data by competing risks. Biometrics 28, 475-488.

Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Klein, J.P. and Basu A.P. (1981). Replacing Dependent Systems by Independent Systems in a Competing Risk Framework with Applications'. Submitted for publication.

Kotz, S. (1974). 'Multivariate Distributions at a Cross Road in Statistical Distributions in Scientific Work VI. Reidell Publishing Co., Boston, 247-270.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35: 139-156.

Langberg, N., Proschan, F. and Quinzi, A.J. (1978). Converting Dependent Models into Independent Ones, Preserving Essential Features Ann. Prob., 6, 174-181.

Langberg, N., Proschan, F. and Quinzi, A.J. (1981). Estimating Dependent Lifelengths with Applications to the Theory of Competing Risks . Ann. Statist. 9:157-167.

Lee L. and Thompson, W.A., Jr. (1974). Results on failure time and pattern for the series system. Pp 291-302 in: Proschan and Serfling (1974).

Mann, N.R. and Grubbs, F.E. (1974). Approximately optimum confidence bounds for system reliability based on component test data. Technometrics 16:335-347.

Mann, N.R., Schafer, R.E., and Singpurwalla, N.D. (1974) Methods for the Statistical Analysis of Reliability and Life Data, Wiley, N.Y.

Marshall, A.W. and Olkin, I. (1967). A Multivariate Exponential Distribution J. Amer. Statist. Assoc. 66, 30-40.

McLean, T.J. (1981). Customer's Risk Evaluation. Paper presented to Modeling and Simulation Conference. Pittsburgh, Pa.

Mendenhall, W. and Lehman, E.H., Jr. (1960). An approximation to the negative moments of the positive binomial useful in life testing. Technometrics 2:227-242.

Miller, D.R. (1977). A note on independence of multivariate lifetimes in competing risks. Ann Statist.5: 576-579.

Moeschberger, M.L. and David, H.A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics 27, 909-933.

Moeschberger, M.L. (1974). Life tests under dependent competing causes of failure. Technometrics 16, 39-47.

Oakes, D. (1982). A model for association in bivariate survival data. J. Roy. Statis. Soc. 44:414-422.

Peterson, A.V., Jr. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci. 73: 11-13.

Peterson, A.V., Jr. (1978). Dependent competing risks: bounds for net survival functions with fixed crude survival functions. Environmental International 1: 351-355.

Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Jr., Flournoy, N., Farewell, V.T., and Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. Biometrics 34: 541-554.

Proschan, F. and Serfling, R.J. (Eds.) (1974). Reliability and Biometry: Statistical Analysis of Lifelength. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Rose, D.M. (1973). Investigation of dependent competing risks. Ph.D. dissertation, University of Washington.

Thompson, W.A., Jr. (1979). Technical note: competing risk presentation of reactor safety studies. Nuclear Safety 20: 414-417.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. 72, 20-22.

Tsiatis, A. (1977). Nonidentifiability problems with the reliability approach to competing risks. Technical Report No. 490, University of Wisconsin-Madison.

Weinberg, Alvin (1978). Editorial. Environmental International 1:285-287.

APPENDIX A

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Comparison of several methods of estimating the survival function when there is extreme right censoring	5. TYPE OF REPORT & PERIOD COVERED Annual 10/1/84 to 10/31/85
	6. PERFORMING ORG. REPORT NUMBER 714837 MPN 763265
7. AUTHOR(s) M.L. Moeschberger and John P. Klein	8. CONTRACT OR GRANT NUMBER(s) AFSOR -82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington D.C. 20332	12. REPORT DATE November 26, 1985
	13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington DC 20332	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited; approved for public release.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Adjusted Kaplan-Meier Survival Estimation, Bias of Survival Function, Life-Testing, Survival Analysis, Right Censoring

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

When there is extreme censoring on the right, the Kaplan-Meier product limit estimator is knownto be a biased estimator of the survival function. Several modifications of the Kaplan-Meier estimator are examined and compared with respect to bias and mean squared error.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BIOMETRICS 41, 253–259
March 1985

A Comparison of Several Methods of Estimating the Survival Function when There Is Extreme Right Censoring

M. L. Moeschberger[1] and John P. Klein[2]

Departments of Preventive Medicine[1] and Statistics[2], The Ohio State University, 320 West 10th Avenue, Columbus, Ohio 43210, U.S.A.

SUMMARY

When there is extreme censoring on the right, the Kaplan–Meier product-limit estimator is known to be a biased estimator of the survival function. Several modifications of the Kaplan–Meier estimator are examined and compared with respect to bias and mean squared error.

1. Introduction

In human and animal survival studies, as well as in life-testing experiments in the physical sciences, one method of estimating the underlying survival distribution (or the reliability of a piece of equipment) which has received widespread attention is the Kaplan–Meier product-limit estimator (Kaplan and Meier, 1958).

For the situation in which the longest time an individual is in a study (or on test) is not a failure time, but rather a censored observation, it is well known that there are many complex problems associated with any statistical analysis (Lagakos, 1979). In particular, the Kaplan–Meier product-limit estimator is biased on the low side (Gross and Clark, 1975). In the case of many censored observations larger than the largest observed failure time, this bias tends to be worse. Estimated mean survival time and selected percentiles, as well as other quantities dependent on knowledge of the tail of the survival function, will also exhibit such biases.

A practical situation which motivates this study is a large-scale animal experiment conducted at the National Center for Toxicological Research (NCTR), in which mice were fed a particular dose of a carcinogen. The goal of the experiment was to assess the effects of the carcinogen on survival and on age-specific tumor incidence. Toward this end, mice were randomly divided into three groups and followed until death or until a prespecified group censoring time (280, 420, or 560 days) was reached, at which time all those still alive in a given group were sacrificed. Often there were many surviving mice in all three groups at the sacrifice times.

In general, we consider an experiment in which n individuals are under study and censoring is permitted. Let $t_{(1)}, \ldots, t_{(m)}$ denote the m ordered failure times of those m individuals whose failure times are actually observed ($t_{(1)} \leq \cdots \leq t_{(m)}$). The remaining $n-m$ individuals have been censored at various points in time. It will be useful to introduce the notation S_j to denote the number of survivors just prior to time $t_{(j)}$; that is, S_j is the number of individuals still under observation at time $t_{(j)}$, including the one that died at $t_{(j)}$. Then the Kaplan–Meier product-limit estimator (assuming no ties among the $t_{(j)}$) of

Key words: Adjusted Kaplan–Meier survival estimation; Bias of survival function; Life-testing; Right censoring; Survival analysis.

the underlying survival function, $\bar{P}(t) = \Pr(T > t)$, is

$$\hat{P}(t) = \begin{cases} 1 & \text{for } t < t_{(1)} \\ \prod_{r=1}^{j} (S_r - 1)/S_r & \text{for } t_{(j)} \leq t < t_{(j+1)} \\ 0 & \text{for } t \geq t_{(m+1)} \end{cases} \quad (1)$$

for $j = 1, \ldots, m$, where $t_{(m+1)} = t_c$ if the longest time an individual is on study is a censoring time or $t_{(m+1)} = \infty$ if the longest time an individual is on study is a death.

This paper first proposes, in §2, some methods of "completing" the Kaplan-Meier estimator of the survival function by (i) replacing those censored observations that are larger than the last observed failure time by their expected order statistics; (ii) using a Weibull distribution to estimate the tail probability $\bar{P}(t)$, for $t > t_c$; and (iii) employing a method suggested by Brown, Hollander, and Korwar (BHK) (1974). The second purpose is to demonstrate the magnitude of the bias and mean squared error (MSE) of the Kaplan-Meier estimator and to compare all methods of "completing" $\hat{P}(t)$ in the context of the aforementioned mouse study, utilizing simulated lifetimes from exponential, Weibull, lognormal, and bathtub-shaped hazard function distributions. These results are presented in §3.

2. Completion of Kaplan-Meier Product-Limit Estimator

2.1 *Expected Order Statistics*

One method of attempting to "complete" $\bar{P}(t)$, $t > t_c$, would be to "estimate" the failure times for those censored observations that are larger than the longest observed lifetime. Let n_c be the number of censored observations larger than $t_{(m)}$. A theorem regarding the conditional distributions of order statistics states that for a random sample of size n from a continuous parent, the conditional distribution of $T_{(u)}$, given $T_{(n-n_c)} = t_{(n-n_c)}$, $u > n - n_c$, is just the distribution of the $(u - n + n_c)$th order statistic in a sample of size n_c drawn from the parent distribution truncated on the left at $t = t_{(n-n_c)}$ (see David, 1981, p. 20).

For computational purposes, take t_c as an estimate of the $(n - n_c)$th order statistic. Then find the expected value of the n_c order statistics from the parent distribution truncated on the left at t_c. Since the Weibull distribution with survival function $\bar{P}(t) = \exp(-t^k/\theta)$ has been widely accepted as providing a satisfactory fit for lifetime data, it seems reasonable to employ the results of Weibull distribution theory to complete $\bar{P}(t)$, $t > t_c$. (It should be noted that any distribution which is reasonable for the specific situation may be used.) The expected values of Weibull order statistics up to sample size 40 for location parameter equal to 1 and shape parameter equal to .5 (0.5)4(1)8 may be found in Harter (1969). For larger sample sizes, he states a recurrence relation which may be used.

To compute expected values of the n_c order statistics in question, values for k and θ must be chosen. One approach is to use the maximum likelihood estimators, $\hat{k}$ and $\hat{\theta}$, computed by using all observations to estimate k and θ. A second approach, due to White (1969), uses least squares estimates of k and θ obtained by fitting the model

$$\ln(t_{(j)}) = (1/k) \ln \theta + (1/k) \ln[H(t_{(j)})] \quad (2)$$

to the $t_{(j)}$'s, where $H(t_{(j)})$ is the estimated cumulative hazard rate at $t_{(j)}$ obtained from the Kaplan-Meier estimator. In our Monte Carlo study, we found the maximum likelihood estimators performed better than the least squares estimators in all cases. Consequently, the method of least squares will be dropped from future discussion in this paper.

The survival function for a Weibull random variable, truncated on the left at t_c, is

$$\bar{P}_T(t) = \exp[-(t^k - t_c^k)/\theta], \qquad t > t_c. \quad (3)$$

So, by the theorem on order statistics stated at the beginning of this section, the conditional distribution of $T_{(u)}$, given $T_{(n-n_c)} = t_{(n-n_c)}$ $(u = n - n_c + 1, \ldots, n)$ will be approximated by the $(u - n + n_c)$th order statistic in a sample of n_c drawn from (3). For simplicity, let $j = u - n + n_c$, so that $j = 1, \ldots, n_c$. Now the expected value of the jth order statistic from (3) is

$$\mathrm{E}(T_{j:n_c}) = n_c\binom{n_c - 1}{j - 1} \int_{t_c}^{\infty} t[P_T(t)]^{j-1}[\bar{P}_T(t)]^{n_c-j+1}(kt^{k-1}/\theta)\, dt$$

$$= n_c\binom{n_c - 1}{j - 1} \int_{0}^{\infty} (y^k + t_c^k)^{1/k}[P(y)]^{j-1}[\bar{P}(y)]^{n_c-j+1}(ky^{k-1}/\theta)\, dy \tag{4}$$

where $\bar{P}(y) = \exp(-y^k/\theta)$, $y = (t^k - t_c^k)^{1/k} \geq 0$ and $T_{j:n_c}$ is the jth order statistic in a sample of size n_c. Equation (4) can also be written as

$$\mathrm{E}(T_{j:n_c}) = n_c\binom{n_c - 1}{j - 1} \int_{0}^{\infty} (\theta z^k + t_c^k)^{1/k}[P(z)]^{j-1}[\bar{P}(z)]^{n_c-j+1}kz^{k-1}\, dz \tag{5}$$

where $\bar{P}(z) = \exp(-z^k)$, $z = (y/\theta)^{1/k} \geq 0$. Now $\mathrm{E}(T_{j:n_c})$ may be crudely estimated by

$$\{\hat{\theta}[\mathrm{E}(Z_{j:n_c})]^{\hat{k}} + t_c^{\hat{k}}\}^{1/\hat{k}} \tag{6}$$

where $\mathrm{E}(Z_{j:n_c})$ is the expected value of the jth order statistic from a sample of size n_c determined from Harter's (1969) tables or recurrence relation, and $\hat{\theta}$ and $\hat{k}$ are maximum likelihood estimators of θ and k, respectively.

These n_c estimated expected order statistics may then be treated as "observed" lifetimes in adjusting (or "completing") the estimated survival function computed in (1). The area under the estimated survival function up to t_c remains unchanged. The area under the extended estimated survival function based on the n_c estimated expected order statistics is then added to the initial area to obtain a more precise estimate of $\bar{P}(t)$ [estimated order statistic (EOS) extension].

2.2 *Weibull Maximum Likelihood Techniques*

A straightforward approach to completing $\hat{\bar{P}}(t)$ is to set

$$\hat{\bar{P}}(t) = \exp(-t^k/\theta) \qquad \text{for} \quad t > t_c. \tag{7}$$

Estimates of k and θ based on all observations can be obtained by either the maximum likelihood (WTAIL) or the least squares method. However, our study found the completion using maximum likelihood estimators was always better in terms of bias and mean squared error.

One suggestion for ostensibly improving this estimator would be to "tie" the estimated tail to the product-limit estimator at t_c. Two methods were attempted to accomplish this goal. First, the likelihood was maximized with respect to k and θ subject to the constraint that $\exp(-t_c^k/\theta) = \hat{\bar{P}}(t_c)$. This method will be referred to as the restricted MLE tail probability estimate (RWTAIL extension). Second, a scale-shift was performed on the tail probability in (7) to tie it to the product-limit estimator. This method led to higher biases and mean squared errors of the survival function and will be dropped from further discussion in this paper.

2.3 *BHK-Type Methods*

The Brown-Hollander-Korwar completion of the product-limit estimator sets

$$\hat{\bar{P}}(t) = \exp(-t/\theta^*) \qquad \text{for} \quad t > t_c \tag{8}$$

where θ^* satisfies $\hat{P}(t_c) = \exp(-t_c/\theta^*)$. In the BHK spirit we tried to complete $\bar{P}(t)$ by a Weibull function which used estimates of k and θ, k^* and θ^*, that satisfied the following two relations:

$$\hat{P}(t_{(m)}) = \exp(-t_{(m)}^{k^*}/\theta^*)$$

and

$$\hat{P}(t_{(m-1)}) = \exp(-t_{(m-1)}^{k^*}/\theta^*).$$

The latter method also led to consistently poor performance and the results will not be presented.

Table 1
Bias/100 (and MSE/100²) for estimating mean survival time for various methods of completion

Distribution	μ	Mean % censored at 560 days	K–M	BHK extension	Estimated order statistic extension	Weibull WTAIL extension	Restricted Weibull RWTAIL extension
Weibull	400	18.7	−2.000[w]	−1.462	−.101[b]	.131	.206
			(4.034)[w]	(2.271)	(1.172)	(1.160)[b]	(1.543)
$k = .5$	500	22.3	−2.802[w]	−2.078	−.176[b]	.208	.299
			(7.886)[w]	(4.498)	(1.922)[b]	(2.344)	(3.292)
	600	25.5	−3.625[w]	−2.704	−.187[b]	.344	.479
			(13.179)[w]	(7.522)	(3.025)[b]	(4.275)	(6.031)
	400	24.6	−.991[w]	−.047	−.046	.016[b]	.0379
			(1.011)[w]	(.215)[b]	(.257)	(.275)	(.343)
$k = 1$	500	32.6	−1.632[w]	−.049	−.047[b]	.073	.116
			(2.696)[w]	(.416)[b]	(.535)	(.508)	(.705)
	600	39.3	−2.359[w]	.022[b]	.034	.140	.214
			(5.592)[w]	(.596)[b]	(.987)	(1.023)	(1.353)
	400	7.5	−.036	.136[w]	−.005	.003[b]	.004
			(.012)[b]	(.053)[w]	(.013)	(.014)	(.014)
$k = 4$	500	34.6	−.314	1.507[w]	−.020	.014[b]	.019
			(.109)	(2.830)[w]	(.036)[b]	(.041)	(.044)
	600	59.9	−.903	5.982[w]	.144	.028[b]	.039
			(.822)	(41.430)[w]	(4.168)	(.147)[b]	(.157)
Lognormal	400	20.6	−.868[w]	−.178[b]	−.544	−.586	−.412
			(.777)[w]	(.179)[b]	(.363)	(.403)	(.267)
$k = 1$	500	29.0	−1.427[w]	−.150[b]	−.865	−.918	−.696
			(2.060)[w]	(.323)[b]	(.855)	(.938)	(.644)
	600	36.9	−2.079[w]	−.022[b]	−1.234	−1.281	−1.038
			(4.345)[w]	(.571)[b]	(1.679)	(1.800)	(1.301)
	400	8.6	−.070	.129[w]	−.047	−.053	−.027[b]
			(.014)[b]	(.056)[w]	(.014)[b]	(.014)[b]	(.014)[b]
$k = 4$	500	29.1	−.330	1.033[w]	−.170	.181	−.135[b]
			(.118)	(1.459)[w]	(.051)	(.055)	(.043)[b]
	600	54.5	−.853	4.430[w]	−.391	−.392	−.356[b]
			(.734)	(23.159)[w]	(.199)	(.199)	(.177)[b]
Bathtub	400	18.6	−1.069	−.185	−.170	1.125[w]	.063[b]
			(1.175)	(.234)[b]	(.260)	(1.745)[w]	(.361)
$p = .1$	500	26.1	−1.722[w]	−.259	−.202	1.523	.046[b]
			(2.996)	(.427)[b]	(.560)	(3.230)[w]	(.608)
	600	32.6	−2.452[w]	−.362	−.310	1.761	.047[b]
			(6.043)[w]	(.727)[b]	(.982)	(4.490)	(1.254)
	400	8.1	−1.786[w]	−1.543	−1.547	−.936	.343[b]
			(3.218)[w]	(2.463)	(2.476)	(1.081)	(.544)[b]
$p = .4$	500	13.3	−2.370[w]	−1.826	−1.814	−.825	.585[b]
			(5.649)[w]	(3.472)	(3.446)	(1.031)[b]	(1.303)
	600	18.7	−3.072[w]	−2.191	−2.175	−.875	.841[b]
			(9.466)[w]	(5.013)	(4.983)	(1.285)[b]	(2.792)

[b] Best estimation method.
[w] Worst estimation method.

3. A Comparison of the Various Methods

A simulation study of data such as that collected at NCTR was performed. Three groups of 48 lifetimes were simulated with all testing stopping at 280, 420, and 560 days, respectively, for the three groups. Distributions with mean survival times of 400, 500, and 600 days were used. The generated lifetimes greater than or equal to the sacrifice time for each particular group were considered as censored. The remaining set of observed lifetimes, along with the number censored at the three sacrifice times, constituted a single sample. For each of the distributions studied, 1000 such samples were generated. Weibull distributions with shape parameters .5, decreasing failure rate, 1, constant failure rate, and 4,

Table 2
Bias/100 (and MSE/100²) for estimating 90th percentile for various methods of completion

Distribution	μ	K–M	BHK extension	Estimated order statistic extension	Weibull WTAIL extension	Restricted Weibull RWTAIL extension
Weibull	400	−5.017^{w}	−2.858	1.691	.234^{b}	.458
		(25.185)^{w}	(9.358)	(16.424)	(7.524)^{b}	(10.812)
$k = .5$	500	−7.655^{w}	−4.620	1.897	.418^{b}	.642
		(58.604)^{w}	(22.711)	(24.276)	(14.319)^{b}	(21.442)
	600	−10.306^{w}	−6.390	2.213	.734^{b}	1.064
		(106.21)^{w}	(42.449)	(36.895)	(25.419)^{b}	(37.911)
	400	−3.610^{w}	.064^{b}	.248	.084	.067
		(13.035)^{w}	(1.892)^{b}	(2.423)	(1.980)	(2.945)
$k = 1$	500	−5.913^{w}	.096^{b}	.289	.121	.306
		(34.963)^{w}	(2.995)^{b}	(4.681)	(4.361)	(5.903)
	600	−8.216^{w}	.244^{b}	.610	.418	.550
		(67.459)^{w}	(4.198)^{b}	(9.247)	(8.331)	(10.792)
	400	−.045	.098^{w}	−.007^{b}	−.037	−.011
		(.038)^{b}	(.236)^{w}	(.060)	(.047)	(.063)
$k = 4$	500	−1.195	5.324^{w}	−.031	−.026	.024^{b}
		(1.429)	(33.091)^{w}	(.146)	(.141)^{b}	(.177)
	600	−2.554	17.913^{w}	.120	.090	.068^{b}
		(6.524)	(355.02)^{w}	(.794)	(.676)	(.641)^{b}
Lognormal	400	−2.628^{w}	−.044^{b}	−1.263	−1.758	−.967
		(6.908)^{w}	(1.526)^{b}	(1.979)	(3.407)	(1.673)
$k = 1$	500	−4.680^{w}	.213^{b}	−2.354	−2.718	−1.908
		(21.902)^{w}	(2.708)^{b}	(6.153)	(7.909)	(4.751)
	600	−6.736^{w}	.759^{b}	−3.507	−3.766	−2.980
		(45.373)^{w}	(4.764)^{b}	(13.123)	(14.981)	(10.257)
	400	−.085	.161	−.038	−.162^{w}	−.024^{b}
		(.060)^{b}	(.409)^{w}	(.081)	(.065)	(.093)
$k = 4$	500	−1.251	3.722^{w}	−.584	−.657	−.484^{b}
		(1.566)	(17.654)^{w}	(.403)	(.495)	(.318)^{b}
	600	−2.621	13.695^{w}	−1.214	−1.236	−1.158^{b}
		(6.872)	(210.30)^{w}	(1.616)	(1.662)	(1.498)^{b}
Bathtub	400	−3.629^{w}	−.177	.053^{b}	−.104	.105
		(13.167)^{w}	(1.717)^{b}	(2.052)	(2.058)	(3.190)
$p = .1$	500	−6.068^{w}	−.457	−.071	−.208	.004^{b}
		(36.826)^{w}	(2.955)^{b}	(4.702)	(3.619)	(5.245)
	600	−7.997^{w}	−.318	.043	−.244	−.014^{b}
		(63.954)^{w}	(4.330)^{b}	(7.786)	(7.608)	(9.923)
	400	−.347	.143^{b}	.276	1.154^{w}	.981
		(.273)^{b}	(.844)	(1.078)	(3.877)	(4.747)^{w}
$p = .4$	500	−1.425	.521^{b}	.764	1.699	1.718^{w}
		(2.035)	(1.540)^{b}	(2.067)	(8.574)	(10.714)^{w}
	600	−3.554^{w}	−.137	.132^{b}	2.304	2.450
		(12.628)^{w}	(1.804)^{b}	(2.352)	(17.530)	(22.456)

^{b} Best estimation method.
^{w} Worst estimation method.

increasing failure rate, were used. Lognormal distributions, failure rate changes from increasing to decreasing, with first two moments comparable to the above Weibull distributions with $k = 1$ and $k = 4$, were also used. Finally, a bathtub hazard model of Glaser (1980), failure rate changes from decreasing to increasing, was used. This distribution is a mixture of an exponential of parameter λ with probability $1 - p$ and a gamma with parameter λ and index 3 with probability p. Mixing parameters of $p = .1$ and $p = .4$ were used.

The bias and MSE for the estimation of the tail probabilities, i.e., the completed portion of the product-limit estimator, were calculated for each hypothesized distribution and for each competing method of completion. Since these results were extremely similar to those found in estimating mean survival time, $\hat{\mu} = \int_0^\infty \hat{P}(t)\, dt$, we show only the bias and MSE of each competing estimator of μ in Table 1. This also allows us to demonstrate the magnitude of the bias and MSE of the product-limit estimator of μ. The bias and MSE for estimating the 90th percentile are also presented for the various estimation methods in Table 2. As one would expect, the Kaplan–Meier (K–M) estimator performs considerably more poorly than the other estimation schemes. The BHK extension does very well if the underlying distribution is exponential or lognormal with first two moments compatible with the exponential. BHK does reasonably well for the bathtub-shaped hazard model, but it performs very poorly for the Weibull with increasing failure rate and for the lognormal with first two moments compatible with the Weibull.

The remaining three extensions (EOS, WTAIL, and RWTAIL) appear to be somewhat comparable. Each of them is best under certain circumstances although many times the biases and MSEs are so close to one another that they are essentially equivalent. Only the EOS extension has the desirable property of never being worst. It usually is competitive with the method that is best. Ordering the extensions from the standpoint of simplicity, from simplest to most complex, we have BHK, WTAIL, RWTAIL, and EOS.

In summary, the Kaplan–Meier estimator should probably be extended in the presence of extreme right censoring. The choice of extension depends on one's knowledge of the distribution of lifetimes under consideration and the extent of computer facilities available. If the data follow an exponential-type distribution or if no computer facilities are present, the BHK method is the extension of choice due to its simplicity. If the data exhibit a nonconstant failure rate and computer facilities are available, then the RWTAIL or EOS extensions seem to be advisable.

Acknowledgement

This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-82-0307.

Résumé

On sait que l'estimateur de Kaplan–Meier est un estimateur biaisé de la fonction de survie quand le pourcentage d'observations censurées est très élevé. Plusieurs modifications de l'estimateur de Kaplan–Meier sont examinées et comparées du point de vue de leurs biais et écarts moyens quadratiques.

References

Brown, B. W., Jr., Hollander, M., and Korwar, R. M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In *Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proschan and R. J. Serfling (eds), 291–302. Philadelphia: Society for Industrial and Applied Mathematics.

David, H. A. (1981). *Order Statistics*. New York: Wiley.

Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association* **75**, 667–672.

Gross, A. J. and Clark, V. A. (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley.

Harter, H. L. (1969). *Order Statistics and Their Use in Testing and Estimation. Vol.* 2. Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics* **35**, 139–156.

White, J. S. (1969). The moments of log-Weibull order statistics. *Technometrics* **11**, 373–386.

Received June 1980; *revised May* 1984.

Appendix B

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) BOUNDS ON NET SURVIVAL PROBABILITIES FOR DEPENDENT COMPETING RISKS		5. TYPE OF REPORT & PERIOD COVERED 10/1/84 to 10/31/85
		6. PERFORMING ORG. REPORT NUMBER 714837 MPN 763264
7. AUTHOR(s) John Klein and M. L. Moeschberger		8. CONTRACT OR GRANT NUMBER(s) AFOSR-82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University 1314 Kinnear Road Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington, D.C. 20332		12. REPORT DATE November 26, 1985
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington, D.C. 20332		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited; approved for public release

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Key words: Competing Risks, Product Limit Estimator, Net Survival Function, Coefficient of Concordance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Improved bounds on the marginal survival function based on data from a competing risk experiment are obtained. These bounds are obtained by specifying a range of possible concordances for the risks. These bounds are tighter than those of Peterson (1976). A comparison to other existing bounds is also made.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BOUNDS ON NET SURVIVAL PROBABILITIES

FOR DEPENDENT COMPETING RISKS

by

John P. Klein
Department of Statistics
The Ohio State University

and

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University

SUMMARY

Improved bounds on the marginal survival function based on data from a competing risk experiment are obtained. These bounds are obtained by specifying a range of possible concordances for the risks. These bounds are tighter than those of Peterson (1976). A comparison to other existing bounds is also made.

Key Words: Competing Risks, Product Limit Estimator, Net Survival Function, Coefficient of Concordance.

I. INTRODUCTION

A common problem in survival analysis is to estimate the marginal survival function of the time, X, until some event such as remission, component failure, or death due to a specific cause occurs. Often observation of this main event of interest is impossible due to the occurrence of a competing risk at some time Y < X, such as censoring, failure of a different component in a series system, or death from some cause not related to the study. Standard statistical methods, which assume these competing risks are independent, estimate the marginal survival function by the Product Limit Estimator of Kaplan and Meier (1958). This estimator has been shown to be consistent for the marginal survival function by Langberg, Proschan and Quinzi (1981) when the risks follow a constant sum model defined by Williams and Lagakos (1977). When the risks are not in the class of constant sum models, the Product Limit Estimator is inconsistent and, in such cases, the investigator may be appreciably misled by assuming independence.

In the competing risks framework we observe $T = \text{minimum}(X,Y)$ and $I = \chi(X \leq Y)$ where $\chi(\cdot)$ denotes the indicator function. Tsiatis (1975) and others have shown that the pair (T,I) provides insufficient information to determine the joint distribution of X and Y. That is, there exists both an independent and a dependent model for (X,Y) which produces the same joint distribution for (T,I). However, these "equivalent" independent and dependent joint distributions may have quite different marginal distributions. Also, due to this identifiability problem, there may be several dependent models with different marginal structures which will yield the same observable information, (T,I). In light of the consequences of the untestable independence assumption in using the Product Limit estimator to estimate the marginal survival function of X, it is important to have bounds on this function based on the observable random variables (T,I) and some assumptions on the joint behavior of X and Y.

Peterson (1976) has obtained general bounds on the marginal survival function of X, $S(x)$, based on the estimable joint distribution of (T,I). Let $Q_1(x) = P(T > x,\ I = 1)$, and $Q_2(x) = P(T > x,\ I = 0)$ be the crude survival functions of T. His bound, obtained from the limits on the joint distribution of (X, Y) obtained by Fréchet (1951), is

$$Q_1(x) + Q_2(x) \leq S(x) \leq Q_1(x) + Q_1(0). \tag{1.1}$$

Since these bounds allow for any dependence structure, they can be very wide and provide little useful information to an investigator.

Fisher and Kanarek (1974) have obtained tighter bounds on $S(x)$ in terms of a dependence measure α. Their model assumes that simultaneous to the occurrence of Y an event occurs which either stretches or contracts the remaining life of X by an amount associated with α. That is, $P(X > x | Y = y < x) = P(X > y + \alpha(x-y) | Y > y + \alpha(x-y))$. A large α, for example, implies that a small survival after censoring is the same as α-times as much survival if censoring was not present. They show that if α is assumed known, then the marginal survival function can be estimated from the observable information. Also these estimates, $\hat{S}_\alpha(x)$, are decreasing in α. For their bounds, the investigator specifies a range of possible values $\alpha_L < \alpha < \alpha_u$ so that $S_{\alpha_u}(x) \leq S(x) \leq S_{\alpha_L}(x)$.

Recently, Slud and Rubenstein (1983), have proposed general bounds. They show that knowledge of the function

$$\rho(x) = \lim_{\delta \to 0} \frac{P(x < X < x + \delta | X > x, Y \leq x)}{P(x < X < x + \delta | X > x, Y > x)}$$

along with the observable information (T,I) is sufficient to uniquely determine the marginal distribution of X. These estimates $\hat{S}_\rho(t)$ are decreasing functions of ρ for fixed x. Their bounds are obtained by specifying a range of possible values $\rho_1(x) \leq \rho(x) \leq \rho_2(x)$ so that if $\rho(x)$ is the true function $\hat{S}_{\rho_2}(x) \leq S(x) \leq \hat{S}_{\rho_1}(x)$.

In this paper we obtain different bounds on the marginal survival function by assuming a particular dependence structure on X and Y. These bounds are functions of the observables (T,I) and a familiar dependence measure, the *concordance probability* between X and Y. In Section 2 we describe this model in detail. In Section 3 we derive the bounds and show consistency when the dependence parameter is known. In section 4 these bounds are compared to those obtained by Peterson, Fisher and Kanarek, and Slud and Rubenstein.

II. THE MODEL

The dependence structure we shall employ to model the joint survival was first introduced by Clayton (1978) to model association in bivariate lifetables and, later, by Oakes (1982) to model bivariate survival data. Let $S(x) = P(X \geq x)$, $R(y) = P(Y \geq y)$, with $S(0) = R(0) = 1$, be the continuous univariate survival functions of the death and censoring times, respectively. For $\theta \geq 1$ define $F(x,y) = P(X > x, Y > y)$ by

$$F(x,y) = [\{\frac{1}{S(x)}\}^{\theta-1} + \{\frac{1}{R(y)}\}^{\theta-1} - 1]^{-1/(\theta-1)} \tag{2.1}$$

This joint distribution has marginals S and R. As $\theta \to 1$, then (2.1) reduces to the joint distribution with independent marginals. For $\theta \to \infty$, $F(x,y) \to \min(S(x), R(y))$ the bivariate distribution with maximal positive association for these marginals. The probability of concordance is $\theta/(\theta + 1)$ so that Kendall's (1962) coefficient of concordance is $\tau = (\theta - 1)/(\theta + 1)$ which spans the range 0 to 1.

This model has a nice physical interpretation in terms of the functions $\lambda(x|Y = y)$ and $\lambda(x|Y > y)$, the hazard functions of X given Y = y and X given Y > y, respectively. From (2.1) one can show that

$$\lambda(x|Y = y) = \theta\lambda(x|Y > y)$$

or

$$P(X > x|Y = y) = [P(X > x|Y > y)]^{\theta}. \tag{2.2}$$

For $\theta > 1$ the hazard rate of survival if censoring occurs at time y is θ times the hazard rate of survival if censoring does not occur at time y. This implies that the hazard rate after censoring occurs is accelerated by a factor of θ over the hazard rate if censoring had not occurred. Also when $\theta = 1$, (2.2), reduces to the condition required by Williams and Lagakos (1977) for a model to be constant sum and hence for the usual product limit estimator of $S(t)$ to be consistent (See Basu and Klein (1982) for details).

Oakes (1982) also shows that (2.1) can be obtained from the following random effects model. Let $S^*(x) = \exp\{-[\frac{1}{S(x)}]^{\theta-1}+1\}$ and let $R^*(y)$ be similarly defined. Let W have a gamma distribution with density $g(w) \alpha\, w^{\frac{1}{\theta-1}-1} e^{-w}$ and conditional on $W = w$ let X,Y be independent with survival functions $\{S^*(x)\}^w$ and $\{R^*(y)\}^w$. Then, unconditionally, X,Y have the joint survival function $F(x,y)$ given by (2.1).

For fixed marginals S and R the joint probability density function, $f(x,y)$, can be shown to be totally positive of order 2 for all $\theta \geq 1$. This implies that (X,Y) are positive quadrant dependent. In particular, one can show that for S,R fixed the family of distributions $\mathcal{F} = \{F(x,y): \theta \geq 1\}$ is increasing positive quadrant dependent in θ as defined by Ahmed, et al. (1979).

III. BOUNDS ON MARGINAL SURVIVAL

Suppose that X and Y have the joint distribution (2.1) and let $T = \min(X,Y)$, then the survival function of T is

$$F(T) = \left\{\left[\frac{1}{S(t)}\right]^{\theta-1} + \left[\frac{1}{R(t)}\right]^{\theta-1} - 1\right]^{-\frac{1}{\theta-1}} \tag{3.1}$$

and the crude density function associated with X,

$q_1(t) = \frac{d}{dt} P(T < t, X < Y)$, is given by

$$q_1(t) = \frac{s(t)}{S^{\theta}(t)} [F(t)]^{\theta}, \tag{3.2}$$

where $s(t) = -dS(t)/dt$.

Now consider the differential equation

$$s(t)/S^{\theta}(t) = q_1(t)/[F(t)]^{\theta} \tag{3.3}$$

and suppose θ is known. Then the solution of (3.3) for S(t) is

$$S_{\theta}(t) = \left[1 + (\theta-1) \int_0^t \frac{q_1(u)}{[F(u)]^{\theta}} du\right]^{-\left(\frac{1}{\theta-1}\right)} \quad \text{if } \theta > 1$$

$$= \exp\left(-\int_0^t \frac{q_1(u)}{F(u)} du\right) \quad \text{if } \theta = 1. \tag{3.4}$$

The functions $F(\cdot)$ and $q_1(\cdot)$ are directly estimable from the data one sees in a competing risks experiment. Let $T_1, \ldots, T_n$ denote the observed test times of n individuals put on test and let I_i, $i = 1, \ldots, n$ be 1 or 0 according to whether the T_i was an observation on X_i or Y_i, respectively.

Define $\hat{F}(t) = \sum_{i=1}^{n} \chi(T_i > t)/n$ and $\hat{Q}_1(t) = \sum_{i=1}^{n} \chi(T_i \le t, I_i = 1)$.

Then if θ is known, a natural estimator of $S_\theta(t)$ is

$$\hat{S}_\theta(t) = [1 + (\theta-1) \int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta}]^{-\frac{1}{(\theta-1)}} \quad \text{if } \theta > 1$$

$$= \exp\left(-\int_0^t \frac{d\hat{Q}_1(u)}{\hat{F}(u)}\right) \quad \text{if } \theta = 1 \tag{3.5}$$

For $\theta = 1$, this estimator is of the form of the hazard rate estimator proposed by Nelson (1972). The estimators (3.5) can be expressed in the following form for computation purposes.

$$\hat{S}_\theta(t) = \exp\left\{-\sum_{T_{(i)} \le t_i, I_{(i)} = 1} \frac{1}{(n-i+1)}\right\} \quad \text{if } \theta = 1$$

$$[1 + (\theta-1)n^{\theta-1} \sum_{T_{(i)} < t, I_{(i)} = 1} \frac{1}{(n-i+1)} \theta]^{-\frac{1}{\theta-1}} \tag{3.6}$$

$$\text{if } \theta > 1$$

where $T_{(1)}, \ldots, T_{(n)}$ are the ordered death times.

For θ known and if the true underlying joint distribution of (X,Y) is of the form (2.1) then $\hat{S}_\theta(t)$ is a consistent estimator of $S(t)$ as shown by the following theorem.

<u>Theorem 1</u>. Let (X,Y) have the form (2.1) with marginals $S(t)$, $R(t)$ respectively. Let $\theta \ge 1$ be known. Then on the set where $S(t) > 0$ we have $\hat{S}_\theta(t) \to S(t)$ a.s.

Proof:

For $\theta = 1$, the result follows by a theorem of Langberg, Proschan and Quinzi (1981). Suppose that $\theta > 1$. Note that $\hat{Q}_1(t) \to Q_1(t)$ a.s. and $\hat{F}(u) \to F(u)$ a.s. by the strong law of large numbers. Since $\hat{S}_\theta(t)$ is a continuous function of $\int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta}$ in the support of $\hat{F}(u)$, it suffices to show

$$\int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta} \to \int_0^t \frac{dQ_1(u)}{[F(u)]^\theta} \quad \text{a.s.}$$

Now, after an integration by parts,

$$\int_0^t \frac{d\hat{Q}_1(u)}{[\hat{F}(u)]^\theta} = \frac{\hat{Q}_1(t)}{[\hat{F}(t)]^\theta} - \int_0^t \hat{Q}_1(u) d\left(\frac{1}{\hat{F}^\theta(u)}\right)$$

$$= \frac{\hat{Q}_1(t)}{[\hat{F}(t)]^\theta} - \int_0^t [\hat{Q}_1(u) - Q_1(u)] d\left(\frac{1}{\hat{F}^\theta(u)}\right) + \int_0^t Q_1(u) d\left(\frac{1}{\hat{F}^\theta(u)}\right)$$

$$= \frac{\hat{Q}_1(t) - Q_1(t)}{[\hat{F}(u)]^\theta} - \int_0^t [\hat{Q}_1(u) - Q_1(u)] d\left(\frac{1}{\hat{F}^\theta(u)}\right)$$

$$+ \int_0^t \frac{dQ_1(u)}{\hat{F}(u)^\theta} \,. \tag{3.7}$$

By the dominated convergence theorem

$$\lim_{n\to\infty} \int_0^t \frac{dQ_1(u)}{[\hat{F}(u)]^\theta} = \int_0^t \frac{dQ_1(u)}{[F(u)]^\theta} \quad \text{a.s.,}$$

$$\lim_{n\to\infty} \frac{\hat{Q}_1(t) - \hat{Q}_1(t)}{[\hat{F}(u)]^\theta} = 0. \text{ a.s.},$$

and

$$\lim_{n\to\infty} \sup \{|\hat{Q}_1(u) - Q_1(u)|\} = 0, \text{ a.s.}$$

Hence, applying the above results to (3.7), the result now follows: //

To obtain bounds on the net survival function based on data from a competing risks experiment, we proceed as follows. First, note that from (3.5) it is true that $\hat{S}_\theta(t)$ is a decreasing function of θ for fixed t. Also, as $\theta \to 1^+$ we have $\hat{S}_\theta(t) \uparrow \exp(-\int_0^t \hat{F}^{-1}(u)d\hat{Q}_1(u))$.

which provides an upper bound. Notice that this upper bound corresponds to an assumption of independence. As $\theta \to \infty$ one can show that $\hat{S}_\theta(t) \downarrow \hat{F}(t)$ which corresponds to Peterson's (1976) lower bound.

In practice the above bounds, with $\theta = 1, \infty$, while shorter than Peterson's bounds, may still be quite wide.

Tighter bounds may be obtained by an investigator specifying a range of possible values for θ. If the sample size is sufficiently large and $\theta_1 \leq \theta \leq \theta_2$, then $\hat{S}_{\theta_2}(t) \leq S(t) \leq \hat{S}_{\theta_1}(t)$. Specifying θ_1, θ_2 is equivalent to specifying a range of values $\tau_1 < \tau < \tau_2$ for the coefficient of concordance τ since $\theta = (1+\tau)/(1-\tau)$. Hence the primary value of $\hat{S}_\theta(t)$ is in putting bounds on S(t) rather than on estimation of S(t).

IV. EXAMPLE AND COMPARISONS

To illustrate the bounds obtained in the previous section, consider the mortality data reported in Hoel (1972). The data was collected on a group of RFM strain male mice who were subjected to a dose of 300 rads of radiation at age 5-6 weeks. There were three competing risks, thymic lymphoma, reticulutum cell sarcoma, and other causes of death. For illustrative purposes we consider reticulum cell sarcoma as the risk of interest.

Table 1 reports the value of $\hat{S}_\theta(t)$ for concordance $\tau = (\theta - 1)/(\theta + 1)$. The value of $\hat{S}_\theta(t)$ at $\tau = 0$ corresponds to Nelson's (1972) hazard rate estimator assuming independence. Peterson's upper and lower bound ($\tau = 1$) are also reported as are Fisher and Kanerek's bounds and the Slud and Rubenstein bounds for several values which reflect a positive association between risks.

From Table 1 we first note that Peterson's bounds are very wide. Substantial improvement is obtained if one assumes a non-negative dependence structure between risks (See Table 2). Further tightening of these bounds is achieved by assuming that τ is in the range 0 to .5 where the width of the boundaries is at most about 50% of that of Peterson's bounds.

Substantial improvement in the general bounds is also obtained by the bounds of Fisher and Kanerek or Slud and Rubenstein. The bounds of Fisher and Kanerek assume a specific censoring pattern and require a specification of a stretching constant α. Without some additional information, such specification may be impossible. Slud and Rubenstein's bounds are for the general dependence structure. Their bounds require the specification of the $\rho(t)$ function. This function is a quantity which is not easily conceptualized by investigators from either a statistical or biological perspective. This makes it questionable whether reasoable upper and lower bounds on $\rho(t)$ can be extracted from one's prior beliefs. The major advantage of the bounds printed in this paper is that they require only the specification of an upper and lower concordance, a measure quite familiar to most investigators and easily explainable to nonstatisticians.

Table 1

BOUNDS ON THE NET SURVIVAL FUNCTION FOR RETICULUM CELL SARCOMA

		Proposed Bounds for T											Fisher and Kanerek for α				Slud and Rubenstein for ρ(t)					
ime	Peterson's Upper Bound	0*	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0**	2.	5.	10.	100.	2	4	6	8	10	20
320	.980	.972	.970	.967	.963	.958	.948	.932	.899	.830	.739	.707	.961	.813	.733	.727	.965	.951	.939	.927	.916	.873
525	.949	.917	.907	.893	.873	.843	.796	.722	.624	.533	.473	.455	.600	.514	.509	.492	.887	.836	.792	.755	.724	.621
600	.848	.708	.666	.619	.562	.498	.436	.385	.351	.332	.322	.322	.393	.367	.361	.357	.608	.485	.419	.383	.363	.340
620	.818	.640	.590	.536	.476	.415	.362	.323	.298	.285	.275	.275	.342	.326	.325	.310	.554	.429	.370	.341	.325	.306
650	.747	.458	.390	.329	.273	.231	.202	.185	.176	.171	.168	.168	.240	.219	.217	.211	.322	.221	.192	.183	.179	.174
675	.707	.343	.271	.216	.175	.147	.132	.123	.119	.117	.117	.117	.179	.163	.162	.162	.213	.142	.128	.124	.123	.122
700	.677	.250	.180	.136	.108	.092	.084	.080	.080	.078	.077	.077	.126	.123	.123	.123	.136	.091	.084	.083	.082	.081
750	.626	.062	.049	.035	.029	.027	.026	.026	.026	.026	.026	.026	.062	.062	.062	.062	.040	.031	.030	.030	.030	.030

* Nelson's Hazard Rate Estimator

** Peterson's Low Bound, Slud-Rubenstein's $\rho(t) = \infty$

Table 2

RELATIVE SIZE OF THE BOUNDS ON NET SURVIVAL
FOR AN ASSUMED DEPENDENCE STRUCTURE
AS COMPARED TO PETERSON'S BOUNDS

Time	$0 \leq \tau \leq 1$	$0 \leq \tau \leq .5$	$0 \leq \tau \leq .7$
350	.9707	.0879	.2674
525	.9352	.2449	.5931
600	.7338	.5171	.6787
620	.6722	.5120	.6298
650	.5009	.4420	.4870
675	.3831	.3576	.3797
700	.2883	.2767	.2833
750	.0600	.0600	.0600

References

1. Ahmed, A. H., Langberg, N.A., Leon, R.V., and Proschan, F. (1979). Partial Ordering of Positive Quadrant Dependence, with Applications. Florida State University Technical Report.

2. Basu, A. P., and Klein, J. P. (1982). Some Recent Results in Competing Risks Theory. Survival Analysis, Crowley and Johnson, eds. Institute of Mathematical Statistics, Hayward, CA, 216-229.

3. Clayton, D. G. (1978). A Model for Association in Bivariate Lifetables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, Biometrika 65, 141-151.

4. Fisher, L., and Kanarek, P. (1974). Presenting Censored Survival Data When Censoring and Survival Times May Not Be Independent. Reliability and Biometry Statistical Analysis of Lifelength, Proschan and Serfling, eds. SIAM, Philadelphia, 303-326.

5. Fréchet, M. Sur les Tableaux de Correlation Dont les Marges Sont Donnees. Annales de l'Université de Lyon. Section A, Series 3, 14, 53-77.

6. Hoel, D. G. (1972). A Representation of Mortality Data by Competing Risks. Biometrics 28, 475-488.

7. Kaplan, E. L., and Meier, P. (1958). Nonparametric Estimation from Incomplete Samples. JASA 53, 457-481.

8. Kendall, M. G. (1962). Rank Correlation Methods, 3rd ed., Griffin, London.

9. Langberg, N., Proschan, F., and Quinzi, A. J. (1981). Estimating Dependent Life Lengths, With Applications to the Theory of Competing Risks. Ann Stat 9, 157-167.

10. Nelson, W. (1972). Theory and Application of Hazard Plotting for Censored Failure Data. Technometrics 14, 945-966.

11. Oakes, D. (1982). A Model for Association in Bivariate Survival Data. JRSSB 44, 412-422.

12. Peterson, A. V. (1976). Bounds for a Joint Distribution Function with Fixed Subdistribution Functions: Applications to Competing Risks, Proc. Nat. Acad Sci 73, 11-13.

13. Slud, E. V. and Rubenstein, L. V. (1983). Depending Competing Risks and Summary Survival Curves. Biometrika, 70, pp. 643-650.

14. Tsiatis, A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks. Proc. Nat. Acad. Sci. 72, 20-22.

15. Williams, J. S., and Lagakos, S. W. (1977). Models for Censored Survival Analysis: Constant Sum and Variable Sum Models. Biometrika 64, 215-224.

APPENDIX C

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Independent or Dependent Competing Risks? Does it Make a Difference	5. TYPE OF REPORT & PERIOD COVERED 10/1/84 to 10/31/85
	6. PERFORMING ORG. REPORT NUMBER 714837/763265
7. AUTHOR(s) Melvin L. Moeschberger and John P. Klein	8. CONTRACT OR GRANT NUMBER(s) Grant No. AFOSR 82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 92
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington, D.C. 20332	12. REPORT DATE November 26, 1985
	13. NUMBER OF PAGES 44
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington, D.C. 20332	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Competing risks; Component life; Modeling series systems; Robustness studies; System reliability; Gumbel bivariate exponential; Downton bivariate exponential; Oakes bivariate exponential.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This article investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow either one of the three Gumbel bivariate exponential models, the Downton bivariate exponential, or the Oakes bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. (continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

INDEPENDENT OR DEPENDENT COMPETING RISKS? DOES IT MAKE A DIFFERENCE

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

M.L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

This article investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow either one of the three Gumbel bivariate exponential models, the Downton bivariate exponential, or the Oakes bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life are studied based on life-test data collected on series systems when the assumption of independence is made erroneously. Systems with two components are studied.

KEY WORDS: Competing risks; Component life; Modeling series systems; Robustness studies; System reliability; Gumbel bivariate exponential; Downton bivariate exponential; Oakes bivariate exponential.

1. Introduction

Consider a system consisting of several components linked in series. For such a system the failure of any one of the components causes the system to fail. In a biological or medical context we can consider the components to be different lethal diseases and/or different reasons for removal from the study. In a clinical trials framework the primary response of interest, death or remission, and censoring can be considered as components of the system. This general formulation has been detailed in the theory of competing risks (cf. David and Moeschberger (1978)).

A common assumption in such a formulation is that the component lifetimes are statistically independent. Several authors have shown that based on data from series systems only, this assumption, by itself, is not testable because there is no way to distinguish between independent or dependent component lifetimes (see Basu (1981), Basu and Klein (1982), Miller (1977), Peterson (1976), etc.). However, several authors (see Lagakos (1979) p. 152 and Easterling (1980) p. 131) have pointed out the need to determine, quantitatively, how far off one might be if an analysis is based on an incorrect assumption of independence.

To study the effects of erroneously assuming independence we shall assume that each of the components have exponentially distributed lifetimes when tested separately and that the property of marginal exponentiality will be preserved even though some dependence may be induced when the components are linked in series. The assumption of exponentially distributed component lifetimes has been made by Mann and Grubbs (1974), when finding confidence bounds on system reliability, Boardman and Kendall (1970), when estimating component lifetimes from system data, and Miyamura (1982), when combining component and sytem data. (See Barlow and Proschan (1975) or Mann, Schaffer, and Singpurwalla (1974) for a more complete review.) We shall model the dependence structure by the three models of Gumbel (1960), a model proposed by Downton (1970), and a model described by Oakes (1982). These models are briefly described in Section 2.

The effects of a departure from the assumption of independent component lifetimes will be addressed for two distinct situations. The first situation arises in modeling the performance of a theoretical series system constructed from two components. Here, based on testing each component separately or on engineering design principles, it is reasonable to assume that the components are exponentially distributed with known parameter values. Based on this information, we wish to predict parameters such as the mean life or reliability of a series system constructed from these components. In Section 3 we describe how these quantities are affected by departures from independence.

The second situation involves making inferences about component lifetime distributions from data collected on series systems. Commonly, data collected on such systems are analyzed by assuming a constant-sum model, of which independence is a special case (compare Williams and Lagakos (1977) and Lagakos and Williams (1978)). In Section 4.1 we study the properties of the maximum likelihood estimators of the component mean life calculated under an erroneous assumption of independent exponential component lifetimes as mentioned above. Because of the wide spread use of the nonparametric estimator proposed by Kaplan-Meier (1958) for the component reliability we study in Section 4.2, its properties, when the marginal reliabilities are exponential and independence is incorrectly assumed.

2. The Models

Consider a two component series system with component life lengths X_1, X_2. Suppose that each X_i has an exponential survival function

(2.1) $\bar{F}_i(t) = P(X_i > t) = \exp(-\lambda_i t),\ \lambda_i > 0,\ t \geq 0.$

This assumption is made on the basis of extensive testing of each component separately or on knowledge of the underlying mechanism of failure.

To examine the effects of a departure from independence we consider five bivariate exponential models, each with marginals equivalent to (2.1). The first three models are due to Gumbel (1960); the last two models are due to Downton (1970) and Oakes (1982).

2.1 Gumbel's Model A

For this model the joint survival function is

(2.2) $$P(X_1 > x_1,\ X_2 > x_2) = \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} x_1 x_2),$$
$$x_1,\ x_2 \geq 0,\ \lambda_1,\ \lambda_2 > 0,\ 0 \leq \lambda_{12} \leq \lambda_1\lambda_2.$$

The correlation between X_1, X_2 is

$$\rho = -\frac{\lambda_1\lambda_2}{\lambda_{12}} \exp(\lambda_1\lambda_2/\lambda_{12})\ E_i(-\lambda_1\lambda_2/\lambda_{12}) - 1,$$

where $E_i(z) = \int_{-z}^{\infty} \frac{\exp(-u)}{u}\, du$ is the integrated logarithm

For this model ρ varies from $-.40365$ to 0 as λ_{12} decreases from $\lambda_1\lambda_2$ to 0. It is never positive. The regression X_2 on X_1 is non-leanear with

$$E(X_{\_}|X_2 = x_2) = (\lambda_1 + \lambda_{12}x_2 - \lambda_{12}/\lambda_2)/(\lambda_1 + \lambda_{12}x_2)^2.$$

2.2 Gumbel's Model B

For this model the joint survival function is

(2.3) $$P(X_1 > x_1,\ X_2 > x_2) = \exp(-\lambda_1 x_1 - \lambda_2 x_2)\{1 + 4\rho(1-\exp(-\lambda_1 x_1))$$
$$(1-\exp(-\lambda_2 x_2))\},\ \lambda_1,\ \lambda_2 > 0,\ x_1,\ x_2 \geq 0,\ -1/4 \leq \rho \leq 1/4.$$

The correlation, ρ, may be positive or negative. The regression of X_2 on X_1 is again nonlinear with

$$E(X_1|X_2 = x_2) = \{1 + 2\rho - 4\rho \exp(-\lambda_2 x_2)\}/\lambda_1.$$

The effects of a departure from independence on modeling system reliability and estimating component reliabilities has been studied in detail in Moeschberger and Klein (1984).

2.3 Gumbel's Model C

For this model the joint survival function is

$$P(X_1 > x_1, X_2 > x_2) = \exp\{-[(\lambda_1 x_1)^m + (\lambda_2 x_2)^m]^{1/m}\},$$

$$\lambda_1, \lambda_2 > 0, m \geq 1, x_1, x_2 > 0.$$

The correlation is

$$\rho = (4 + 2m) \int_0^{\pi/2} \left[\frac{(\cos\theta \sin\theta)^m}{(\cos\theta^m + \sin\theta^m)^{2+2/m}} \right] d\theta - 1$$

which varies from 0 to 1. For this model m = 1 corresponds to independence and as $m \to \infty$

(2.5) $P(X_1 > x_1, X_2 > x_2) \to \text{minimum}\,[\exp(-\lambda_1 x_1), \exp(-\lambda_2 x_2)]$

the Fréchet (1958) upper bound for these marginals.

2.4 Downton's Model

Downton (1970) suggests modeling bivariate exponential systems by a successive damage model. This model assumes that in a two component system the times between successive shocks on each component have independent exponential distributions and that the number of shocks required to cause each component to fail follows a bivariate geometric distribution. The joint probability density function of the component life times is

(2.5) $$f(x_1,x_2) = \frac{\lambda_1\lambda_2}{1-\rho} \exp\left(-\frac{(\lambda_1 x_1 + \lambda_2 x_2)}{1-\rho}\right) I_0\left(\frac{2\sqrt{\rho\lambda_1\lambda_2 x_1 x_2}}{1-\rho}\right),$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero, and $\lambda_1, \lambda_2 > 0$, $x_1, x_2 \geq 0$, $0 \leq \rho \leq 1$. The correlation between X_1, X_2 is ρ which spans the interval [0,1]. As $\rho \to 1$ the joint survival function of X_1, X_2 approaches the upper Fréchet distribution (2.5). For this model
$E(X_1 | X_2 = x_2) = (1-\rho/\lambda_1 + \rho\ \lambda_2 x_2/\lambda_1$.

2.5 Oakes' Model

Oakes (1982) has proposed a model for bivariate survival data. This model was first proposed by Clayton (1978) to model association in bivariate lifetables. Special cases of Oakes' general model have been suggested by Lindley and Singpurwalla (1985) and Hutchinson (1981).

For this model the joint survival probability is

$$(2.7)\quad P(X_1 > x_1, X_2 > x_2) = [\exp(\lambda_1(\theta-1)x_1) + \exp(\lambda_2(\theta-1)x_2)-1]^{-1/\theta-1)}$$

where $\lambda_1, \lambda_2 > 0$, $\theta \geq 1$, $x_1, x_2 \geq 0$.

For $\theta = 1$, X_1, X_2 are independent and $P(X_1 > x_1, X_2 > x_2) \to (2.5)$ as $\theta \to \infty$. For this model Kendall's (1962) coeffcient of concordance is $\tau = (\theta-1)/(\theta+1)$ which spans the range 0 to 1. The correlation, ρ, also spans the range 0 to 1 and is found numerically.

This model has the following physical interpretation. Let $r(x_1|X_2 = x_2)$ and $r(x_1|X_2 > x_2)$ be the conditional failure rates of X_1 given $X_2 = x_2$ and $X_2 > x_2$. Then $r(x_1|X_2 = x_2) = \theta\ r(x_1|X_2 > x_2)$.

The model can also be derived from a random effects model. This formulation assumes that when the components are tested separately under ideal conditions the component survival functions are $S_i(t) = \exp[-\exp(\lambda_i t(\theta-1)) + 1]$, $i = 1, 2$, and that when the two components are put in a series system in the operating environment there is a random factor W which simultaneously changes each component life distribution to $S_i^W(t)$. If W has a gamma distribution with density function $f(x) \propto w^{(\theta-1)-1} e^{-w}$ then, unconditionally, the joint survival function (2.7) holds.

2.6 Fréchet Bounds

Fréchet (1958) obtained bounds of the joint survival functions which can be obtained for any set of marginal distributions. For exponential marginals these are

$$\text{MAXIMUM}\,(e^{-\lambda_1 x_1} + e^{-\lambda_2 x_2} - 1,\ 0) \leq P(X_1 > x_1,\ X_2 > x_2) \leq \text{MINIMUM}\,(e^{-\lambda_1 x_1},\ e^{-\lambda_2 x_2})\ .$$

For this set of marginals the lower Fréchet distribution has correlation -.694 and the upper Fréchet distribution has correlation 1.0. These are the minimal and maximal correlations for exponential marginals.

3. Errors in Modeling System Life

Suppose that based on extensive testing or based on theoretical considerations each of the two components in a series system is known to have an exponential distribution, (2.1) with marginal means $1/\lambda_1$, $1/\lambda_2$, respectively. It is of interest to predict the system reliability $\overline{F}(t) = P(X_1 > t,\ X_2 > t)$ and the system mean life $\mu = \int_0^\infty \overline{F}(t)\,dt$. If the investigator assumes that the two components are independent then the system reliability is

(3.1) $\overline{F}_I(t) = \exp(-(\lambda_1 + \lambda_2)t)$ and system mean life is $\mu_I = 1/(\lambda_1 + \lambda_2)$.

If the components are not independent, but in fact follow one of the models in Section 2, then a measure of the effects of incorrectly assuming independence $\Delta(t) = (\overline{F}(t) - \overline{F}_I(t))/\overline{F}_I(t)$ and $\delta = (\mu - \mu_I)/\mu_I$, for predicting system reliability and system mean life, respectively, where $\overline{F}(t)$ and μ are computed under the appropriate dependent model. Values of $\overline{F}(t)$ can be computed directly from (2.2), (2.3), (2.4), (2.6) (by numerical integration) or from (2.7). Expressions for μ are given in Appendix 1. All expressions for $\Delta(t)$ and δ depend on the values of λ_1 and λ_2 only through the ratio $\lambda_1/\lambda_2 = K$ and for $K < 1$ the values are equivalent to those for $K^* = \frac{1}{K}$. For the upper Frechet distribution,

$$\Delta(t_p) = p^{-\frac{K}{K+1}} - 1 \text{ where } K \geq 1 \text{ and } tp = \text{the point}$$

where $\bar{F}_I(t_p) = p$. Also $\delta = 1/K$ for $K > 1$. For the lower Frechet distribution

$$\Delta(t_p) = \begin{cases} p^{-\frac{1}{K+1}} + p^{-\frac{K}{K+1}} - p^{-1} - 1 & \text{if } p^{\frac{K}{K+1}} + p^{\frac{1}{K}} - 1 > 0 \\ -1 & \text{otherwise} \end{cases}$$

and $\delta = \frac{K^2 + K + 1}{K} - \frac{(K+1)^2}{K} Y + (K+1)\, \ln(Y)$ where Y is the solution of the equality $X^K + X = 1$. Table 1 gives the values of $\Delta(t_p) \times 100\%$ and $\delta \times 100\%$ for p = .9, .7, .5, .3, .1 for the upper and lower Fréchet distributions.

From Table 1 we see that the largest percent error occurs when the parameters are equal. Also for fixed K there is relatively small error in estimating system reliability by modeling a dependent system by an independent system when $\bar{F}(t)$ is large. For smaller values of system reliability one can be appreciably misled. Errors in estimating system mean life appear to be substantial unless one component has considerably longer marginal life than the second one. In that instance, one can see instinctively that the correlation would have a minimal impact.

Figures 1A-1C and 2A-2C are plots of $\Delta(tp)$ for p = .24, .5, .75, $\lambda_1 = 1$ and $\lambda_2 = 1, 1.5$ for the five models described in Section 2. Figures 3A, 3B are plots of δ for all five models as a function of the correlation. From these plots note that for positive correlations the most modeling error occurs for the Gumbel C model. For relatively small correlation, $-.25 \leq \rho \leq .25$ there may still be a moderate modeling error, on the order at least ± 10% for predicting system reliability at $\overline{F}_I(t) = .25$, or $\overline{F}_I(t) = .5$ and for estimating the mean system life.

TABLE 1

UPPER AND LOWER BOUNDS ON THE PERCENT ERROR IN MODELING SYSTEM LIFE

	$\bar{F}(T)=0.9$		$\bar{F}(T)=0.7$		$\bar{F}(T)=0.5$		$\bar{F}(T)=0.3$		$\bar{F}(T)=0.1$		MEAN LIFE	
K	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND
1	-0.29	5.41	-3.81	19.52	-17.16	41.42	-100.00	82.57	-100.00	216.23	-38.63	100.00
2	-0.26	3.57	-3.39	12.62	-15.27	25.99	-100.00	49.38	-100.00	115.44	-37.07	50.00
3	-0.22	2.67	-2.86	9.33	-12.90	18.92	-51.52	35.12	-100.00	77.83	-34.85	33.33
4	-0.19	2.13	-2.44	7.39	-11.02	14.87	-44.11	27.23	-100.00	58.49	-32.83	25.00
5	-0.16	1.77	-2.12	6.12	-9.57	12.25	-38.38	22.22	-100.00	46.78	-31.06	20.00
6	-0.14	1.52	-1.87	5.23	-8.45	10.41	-33.91	18.77	-100.00	38.95	-29.53	16.67
7	-0.13	1.33	-1.67	4.56	-7.55	9.05	-30.33	16.24	-100.00	33.35	-28.18	14.29
8	-0.12	1.18	-1.51	4.04	-6.82	8.01	-27.42	14.31	-100.00	29.15	-26.99	12.50
9	-0.11	1.06	-1.37	3.63	-6.22	7.18	-25.02	12.79	-100.00	25.89	-25.93	11.11
10	-0.10	0.96	-1.26	3.30	-5.71	6.50	-22.99	11.57	-100.00	23.28	-24.98	10.00
11	-0.09	0.88	-1.17	3.02	-5.28	5.95	-21.27	10.55	-100.00	21.15	-24.12	9.09
12	-0.08	0.81	-1.08	2.78	-4.91	5.48	-19.78	9.70	-100.00	19.38	-23.34	8.33
13	-0.08	0.76	-1.01	2.58	-4.59	5.08	-18.49	8.98	-100.00	17.88	-22.62	7.69
14	-0.07	0.70	-0.95	2.41	-4.30	4.73	-17.35	8.36	-100.00	16.59	-21.96	7.14
15	-0.07	0.66	-0.90	2.25	-4.05	4.43	-16.35	7.82	-100.00	15.48	-21.35	6.67
16	-0.06	0.62	-0.85	2.12	-3.83	4.16	-15.45	7.34	-100.00	14.50	-20.78	6.25
17	-0.06	0.59	-0.80	2.00	-3.63	3.93	-14.65	6.92	-100.00	13.65	-20.26	5.88
18	-0.06	0.56	-0.76	1.89	-3.45	3.72	-13.93	6.54	-100.00	12.88	-19.76	5.56
19	-0.06	0.53	-0.73	1.80	-3.29	3.53	-13.27	6.20	-100.00	12.20	-19.30	5.26
20	-0.05	0.50	-0.69	1.71	-3.14	3.36	-12.67	5.90	-92.26	11.59	-18.87	5.00
21	-0.05	0.48	-0.66	1.63	-3.00	3.20	-12.13	5.63	-88.34	11.03	-18.46	4.76
22	-0.05	0.46	-0.64	1.56	-2.88	3.06	-11.63	5.37	-84.73	10.53	-18.08	4.55
23	-0.05	0.44	-0.61	1.50	-2.76	2.93	-11.16	5.14	-81.41	10.07	-17.71	4.35
24	-0.04	0.42	-0.59	1.44	-2.66	2.81	-10.74	4.93	-78.34	9.65	-17.37	4.17
25	-0.04	0.41	-0.57	1.38	-2.56	2.70	-10.34	4.74	-75.49	9.26	-17.04	4.00

4. Errors in Estimating Component Parameters

4.1 Parametric Estimation

In this section we examine the effects of incorrectly assuming independence on the magnitude of the estimation error in estimating the first component mean life based on data from series systems. Suppose that n series systems are put on test. For each system we observe the system failure time and which component caused the failure. Let n_i dentoe the number of systems where the system failure was caused by failure of the i^{th} component, $i = 1, 2$, and let T be the total time on test for all n systems. If we assume that the component lifetimes are independent and exponentially distributed then Moeschberger and David (1971) show that the maximum likelihood estimator of μ_1, the first component mean life is

(4.1) $$\hat{\mu}_1 = T/n_1 \text{ for } n_1 > 0.$$

This estimator is asymptotically unbiased and for n finite $E(\hat{\mu}_1) = E(T)\cdot E(1/n_1|n_1 > 0)$ due to the independence of T and n_1.

Suppose now that the two component lifetimes are not independent but follow one of the models discribed in Section 2. If we incorrectly assume independence then a measure of the excess bias due to incorrectly assuming independence is $B = [E(\hat{\mu}_1|\text{Dependent model}) - E(\hat{\mu}_1|\text{independence})]/\mu_1$. For each of the dependent models under consideration T and n_1 are independent. For large n, B converges to $(\mu/p - \mu_1)/\mu_1$ where μ is the mean system life and p is the probability the first component fails first, computed under the dependent model. For finite n, $E(\hat{\mu}_1) = n\,\mu\,E_p(1/n_1|n_1 > 0)$ computed under the appropriate model, where $E_p(1/n_1|n_1 > 0) = \sum_{K=1}^{n} \binom{n}{K} p^K (1-p)^{n-K}/K \,/\, (1-(1-p)^n)$. Expressions for μ and p are given in Appendix 1 and Appendix 2, respectively.

The expressions depend on λ_1, λ_2 only through the ratio $K = \lambda_1/\lambda_2$. For all models, $p = 1/2$ when $K = 1$.

For the upper Fréchet distribution $p = 0$ if $K < 1$; $1/2$ if $K = 1$; and 1 if $K > 1$. Hence for $K < 1$ no failures from the first component are ever observed so that the modeling error B becomes infinite for all n. For $K > 1$, $p = 1$ and $\mu = \mu_1$ so that $B = (1 - E(\hat{\mu}_1 | \text{Independence})/\mu_1)$ which tends to 0 as $n \to \infty$. In this case the models with correlation ranging from 0 to 1 have B increasing for $p < p_0$ and decreasing for $p > p_0$. For the lower Fréchet distribution, p is the value of X which solves the equation $X + X^K - 1 = 0$. For $K < 1$ we have $p < 1/2$ and for $K > 1$ we have $p > 1/2$. Table 2 gives the value of B for $n = 25, 50, \infty$ for the two Fréchet distributions. It also gives the maximum modeling error for the Gumbel C model which is an indication of maximal excess modeling error.

From Table 2 we note that the dependence structure exerts a large effect on estimating the smaller of the two component means and that either effect is most exaggerated for small sample sizes. For $K \geq 1$ there is very little sample size effect on the modeling error. For K strictly bigger than one the maximum bias under the Gumbel C model decreases with K and the correlation at which this maximum is attained also decreases to 0. Figures 4A-4C are plots of B as $n \to \infty$ for K = 3/2, 1, 2/3, respecitvely. Figures 5A-5C are plots of B for n = 10 for K = 3/2, 1, 2/3, respectively.

TABLE 2

RELATIVE MODELING ERROR IN ESTIMATING THE MEAN

	N=20				N=50				N=INFINITY			
	LOWER		GUMBEL C		LOWER		GUMBEL C		LOWER		GUMBEL C	
K	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS
1/10	-167.46	******	1.000	******	-80.09	******	1.000	******	-58.64	******	1.000	******
1/ 9	-142.98	******	1.000	******	-76.35	******	1.000	******	-57.84	******	1.000	******
1/ 8	-123.37	******	1.000	******	-76.68	******	1.000	******	-56.93	******	1.000	******
1/ 7	-107.25	******	1.000	******	-69.03	******	1.000	******	-55.87	******	1.000	******
1/ 6	-93.66	******	1.000	******	-65.34	******	1.000	******	-54.63	******	1.000	******
1/ 5	-81.91	******	1.000	******	-61.54	******	1.000	******	-56.13	******	1.000	******
1/ 4	-71.45	******	1.000	******	-57.50	******	1.000	******	-51.24	******	1.000	******
1/ 3	-61.78	******	1.000	******	-53.00	******	1.000	******	-48.73	******	1.000	******
1/ 2	-52.16	******	1.000	******	-47.31	******	1.000	******	-48.08	******	1.000	******
1	-40.90	105.99	1.000	105.99	-39.45	102.13	1.000	102.13	-38.63	100.00	1.000	100.00
2	-32.56	-2.79	0.510	10.82	-32.84	-1.04	0.518	11.50	-32.12	0.00	0.528	11.91
3	-28.28	-1.82	0.400	5.33	-28.39	-0.69	0.410	5.84	-28.39	0.00	0.424	6.15
4	-25.63	-1.35	0.346	3.39	-25.76	-0.52	0.362	3.76	-25.83	0.00	0.373	4.01
5	-23.62	-1.08	0.312	2.39	-23.80	-0.41	0.330	2.71	-23.90	0.00	0.341	2.91
6	-22.05	-0.89	0.288	1.82	-22.25	-0.34	0.307	2.0[illegible]	-22.36	0.00	0.319	2.26
7	-20.77	-0.76	0.271	1.46	-20.98	-0.29	0.290	1.[illegible]4	-21.11	0.00	0.303	1.83
8	-19.69	-0.67	0.257	1.20	-19.92	-0.26	0.277	1.40	-20.04	0.00	0.289	1.53
9	-18.78	-0.59	0.246	1.01	-19.00	-0.23	0.266	1.19	-19.13	0.00	0.279	1.31
10	-17.98	-0.53	0.237	0.87	-18.20	-0.20	0.257	1.03	-18.33	0.00	0.271	1.14

TABLE 3

ASYMPOTIC BIAS OF THE PRODUCT LIMIT ESTIMATOR

	$\bar{F}(T)=0.7$				$\bar{F}(T)=0.5$				$\bar{F}(T)=0.3$			
	LOWER		GUMBEL C		LOWER		GUMBEL C		LOWER		GUMBEL C	
K	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS	BOUND	RHO=1	RHO	BIAS
1/ 9	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 8	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 7	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 6	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 5	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 4	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-100.00	233.33	1.000	233.33
1/ 3	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-52.70	233.33	1.000	233.33
1/ 2	-100.00	42.86	1.000	42.86	-100.00	100.00	1.000	100.00	-24.21	233.33	1.000	233.33
1	-100.00	19.52	1.000	19.52	-100.00	41.42	1.000	41.42	-9.65	82.57	1.000	82.57
2	-100.00	0.00	0.528	3.87	-24.67	0.00	0.528	7.65	-4.40	0.00	0.528	13.67
3	-100.00	0.00	0.424	2.09	-14.98	0.00	0.424	4.10	-2.85	0.00	0.424	7.22
4	-63.03	0.00	0.373	1.38	-10.78	0.00	0.373	2.70	-2.11	0.00	0.373	4.74
5	-1.67	0.00	0.341	1.01	-8.42	0.00	0.341	1.98	-44.50	0.00	0.341	3.47
6	-1.39	0.00	0.319	0.79	-6.91	0.00	0.319	1.54	-34.83	0.00	0.319	2.70
7	-1.18	0.00	0.303	0.64	-5.86	0.00	0.303	1.26	-28.71	0.00	0.303	2.19
8	-1.03	0.00	0.290	0.54	-5.09	0.00	0.290	1.05	-24.45	0.00	0.290	1.83
9	-0.92	0.00	0.279	0.46	-4.50	0.00	0.279	0.90	-21.31	0.00	0.279	1.57
10	-0.82	0.00	0.271	0.40	-4.03	0.00	0.271	0.78	-18.89	0.00	0.271	1.37

Figures 5A-5C for p = .25, 6A-6C for p = .5 and 7A-7C for p = .75 are plots of $\Delta_1(p)$ for the 5 models and k = 3/2, 1, 2/3. As in the previous figures one can see that for even a small departure from independence the relative effect of dependence can be quite large.

4.2 Nonparametric Estimation

A second approach to the problem of estimating component parameters is via the nonparametric estimator of Kaplan and Meier (1958). Investigators who routinely use nonparametric techniques may take this approach in hopes of obtaining estimators that are robust with respect to the assumption of exponentiality. However this estimator is not necessarily robust to the assumption of independence.

The product limit estimator, assuming independent risks is constructed as follows. Suppose that n systems are put on test and let $r_{i1}, \ldots, r_{in_i}$ be the ranks of the ordered n_i failures from cause i, $X_{i(1)}, \ldots, X_{i(n_i)}$, among all n order lifetimes. The estimator of the component reliability for the i^{th} component is

4.2.1
$$\hat{S}_i(x) = \begin{cases} 1 \text{ if} & x < x_{i(1)}, \\ \prod_{j=1}^{j(i,x)} \dfrac{n - r_{ij}}{n - r_{ij} + 1} & x > x_{(i(1)} \end{cases}$$

where $j(i,x)$ is the largest value of j for which $x_{i(j)} < x$. This estimator is asympotically unbiased when the component lifetimes are independent.

When the risks are dependent Klein and Moeschberger (1983) show that $\hat{S}_i(t)$ is not estimating the marginal component reliability, but rather it is estimating consistently another survival function

4.2.2
$$\bar{H}_i(x) = \exp\left(-\int_0^x \frac{d\, Q_i(t)}{\bar{F}(t)}\right) \quad \text{where}$$

$\bar{F}(t) = P(\text{minimum } (X_1,X_2) > t)$ and $Q_i(t) = P(\min(X_1,X_2) \leq t, \min(X_1,X_2) = X_i)$, $i = 1, 2$. Expressions for

$\bar{H}_1(t)$ for the five models of interest are given in Appendix 3.

A measure of the affect of dependence in using the Product limit estimator with dependent risks is $\Delta_1(p) = (\bar{H}_1(t_p) - p)/p$ where t_p is the time where the true component reliability is p. $\Delta_1(p)$ is again only a function of $k = \lambda_1/\lambda_2$. For the upper Fréchet distribution

$$\Delta_1(p) = \begin{cases} p^{-1} - 1 & \text{for} \quad k < 1 \\ p^{-1/2} - 1 & \text{for} \quad k = 1 \\ 0 & \text{for} \quad k > 1 \end{cases} \quad \text{since}$$

for $k < 1$ $\bar{H}_1(t) = 1$ for all t since the first component never fails, while for $k > 1$ all failures are due to the first component. For those models with correlation spanning the range $[0 - 1]$, $\Delta_1(p)$ is increasing for correlations less than ρ^* and decreasing for correlations greater than ρ^* when $k > 1$. For the lower Frechet distribution

$$\bar{H}_1(t_p) = \begin{cases} \exp - \int_p^1 \frac{1}{u + u^k - 1}\, du & \text{for } p \geq (1-Y) \\ 0 & \text{otherwise.} \end{cases}$$

Table 3 shows the value of $\Delta_1(p) \times 100\%$ for p = .7, .5, .3 for the two Frechet distributions. For $k > 1$, the maximum value under the Gumbel C model is also given. As in the parametric estimation problem the largest errors are incurred when $k < 1$. In all cases the effect of a departure from independence is the largest when p is small (i.e. for large t). The effect decreases as k increases reflecting the fact that when $k\lambda_1 >> \lambda_2$ the majority of the system failures are due to the failure of the first component.

5. Conclusions

The results presented in this paper show that for all five bivariate exponential models one may be appreciably misled by falsely assuming independence of component lifetimes in a series system. The amount of error incurred in modeling system reliability not only depends upon the correlation between component lifetimes but also on the level of system reliability. The error in modeling mean system life similarly depends upon the correlation and the length of mean system life. Both quantities depend on the relative magnitudes of the parameters.

For the dual problem of estimating component reliability based on data from a series system, it appears that departures from independence are of greater consequence. Both parametric and nonparametric estimators of relevant component parameters are inconsistent. Bias increases dramatically as the correlation gets further from zero. However, the five models do not exhibit appreciable differences in bias and mean squared error as correlation changes. This suggests that these models may belong to a large class of bivariate exponential distributions which possesses the properties exhibited here.

ACKNOWLEDGEMENT

Research sponsored by the Air Force Office of Scientific Research under contract AFOSR-82-0307.

REFERENCES

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt Rinehart, and Winston.

Basu, A.P. (1981). Identifiability problems in the theory of competing and complementary risks - a survey. Statistical Distribution in Scientific Work (Taillie, Patil, and Baldesaari, Eds.), Dorerecht, Holland: Reidel Publishing Co., 335-348.

Basu, A.P., and Klein, J.P. (1982). Some Recent Results in Competing Risks Theory. Survival Analysis, Crowley and Johnson, eds. Institute of Mathematical Statistics, Hayward, CA, 216-229.

Birnbaum, Z.W. (1979). On the Mathematics of Competing Risks. U.S. Dept. of HEW. Publication No. 79-1351.

Boardman, T.J. and Kendell, B.J. (1970). Estimation in compound exponential failure models. Technometrics 12, 891-900.

Clayton, D.G. (1978). A Model for Association in Bivariate Life-tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, Biometrika 65, 141-151.

David, H.A. and Moeschberger, M.L. (1978). The Theory of Competing Risks. Griffin, London.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. J. Roy. Statist. Soc. B. 32: 408-417.

Easterling, R.G. (1980). Book review of Z.W. Birnhaum monograph. Technometrics 22, 131-132.

Fréchet, M. Sur les Tableaux de Correlation Dont les Marges Sont Donnees. Annales de l'Université de Lyon. Section A, Series 3, 14, 53-77.

Gumbel, E.J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55, 698-707.

Hutchinson, T.P. (1981). Compound Gamma Bivariate Distributions. Metrika, 28, 263-271.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Klein, J.P. and Moeschberger, M.L. (1984). Asymptotic bias of the product limit estimator under dependent competing risks. IAPQR Transactions, 9, 1-7.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35, 139-156.

Lindley, D. and Singpurwalla, N. (1984). Multivariate distributions for the reliability of a system of components sharing a common environment. Presented at the International Conference on Reliability and Quality Control.

Mann, N.R. and Grubbs, F.E. (1974). Approximately optimum confidence bounds for system reliability based on component test data. Technometrics 16, 335-347.

Mann, N.R., Schafer, R.E., and Singpurwalla, N.D. (1974). Methods for the Statistical Analysis of Reliability and Life Data. Wiley, New York.

Miyamura, T. (1982). Estimating component failure rates from combined component and systems data: exponentially distributed component lifetimes. Technometrics 24: 313-318.

Miller, D.R. (1977). A note on independence of multivariate lifetimes in competing risks. Ann. Statist. 5: 576-579.

Moeschberger, M.L. and David, H.A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrika 27, 909-933.

Moeschberger, M.L. and Klein, J.P. (1984). Consequences of departures from independence in exponential series systems. Technometrics 26: 277-284.

Oakes, D. (1982). A model for association in bivariate survival data. J. Roy. Statist. Soc. B. 44: 414-422.

Peterson, A.V., Jr. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci. 73: 11-13.

Williams, J.S. and Lagakos, S.W. (1977). Models for censored survival analysis: Constnat-sum and variable-sum models. Biometrika 64: 215-224.

Appendix 1

Formulas for expected System Life,

$$\text{Gumbel A} - \exp \frac{(\lambda_1+\lambda_2)^2}{4\lambda_1\lambda_2} \sqrt{\frac{\pi}{\lambda_{12}}}\ \bar{\Phi}\,\frac{\lambda_1+\lambda_2}{\sqrt{2}\,\lambda_{12}} \quad , \qquad \text{(A1.1)}$$

Where $\bar{\Phi}(\cdot)$ is the survival function of a standard normal random variable.

$$\text{Gumbel B} - \frac{1}{(\lambda_1+\lambda_2)} + \frac{6\,\rho\,\lambda_1\lambda_2}{(\lambda_1+\lambda_2)(2\lambda_1+\lambda_2)(\lambda_1+2\lambda_2)} \qquad \text{(A1.2)}$$

$$\text{Gumbel C} - (\lambda_1^m+\lambda_2^m)^{-1/m} \qquad \text{(A1.3)}$$

$$\text{Downton} - \frac{(\lambda_1+\lambda_2)(1-\rho)}{[(\lambda_1+\lambda_2)^2 - 4\rho\,\lambda_1\lambda_2]} \qquad \text{(A1.4)}$$

$$+ \frac{[(\lambda_1+\lambda_2) - \sqrt{(\lambda_1+\lambda_2)^2 - 4\rho\,\lambda_1\lambda_2}\,][\lambda_1^2+\lambda_2^2 - 2\rho\,\lambda_1\lambda_2]}{2\lambda_1\lambda_2[(\lambda_1+\lambda_2)^2 - 4\rho\,\lambda_1\lambda_2]}$$

Oakes - found by numerical integration

Appendix 2 - Formulas for $p = P(X_1 < X_2)$

Gumbel A - $P(X_1 < X_2) = 1/2 + (\lambda_1 - \lambda_2)\sqrt{\lambda \frac{\pi}{12}}\exp\left(\frac{(\lambda_1+\lambda_2)^2}{4\lambda_{12}}\right)\overline{\Phi}\left(\frac{\lambda_1+\lambda_2}{2\lambda_{12}}\right),\ \lambda_{12} > 0$

where $\overline{\Phi}(\cdot)$ is the survival function of a standard normal random variable. (A2.1)

Gumbel B - $P(X_1 < X_2) = \frac{\lambda_1}{(\lambda_1+\lambda_2)} + 4\rho\,\frac{\lambda_2(\lambda_1-\lambda_2)}{(\lambda_1+\lambda_2)(\lambda_1+2\lambda_2)(2\lambda_1+\lambda_2)}$

(A2.2)

Gumbel C - $P(X_1 < X_2) = \frac{\lambda_1^m}{(\lambda_1^m+\lambda_2^m)}$ (A2.3)

Downton - $P(X_1 < X_2) = \frac{2\lambda_1\lambda_2(1-\rho)}{\left(\sqrt{(\lambda_1+\lambda_2)^2 - 4\rho\,\lambda_1\lambda_2}\right)\left(\lambda_1-\lambda_2 + \sqrt{(\lambda_1+\lambda_2)^2 - 4\rho\lambda_1\lambda_2}\right)}$

(A2.4)

[illegible] - $P(X_1 < X_2)$ found numerically.

Appendix 3 -

Gumbel A - $\bar{H}_1(x) = \exp(-\lambda_1 x - \lambda_{12} x^2)$ (A3.1)

Gumbel B - $$\bar{H}_1(x) = \exp\left\{-\int_0^x \frac{[1 + 4\rho(1-\exp(-\lambda_2 t))(1-2\exp(-\lambda_1 t))]}{[1 + 4\rho(1-\exp(-\lambda_2 t))(1-\exp(-\lambda_1 t))\ ]} dt\right\} \quad \text{(A3.2)}$$

Gumbel C - $$H_1(x) = \exp\left\{-\frac{\lambda_1^m x}{(\lambda_1^m t + \lambda_2^m)^{\frac{m-1}{m}}}\right\} \quad \text{(A3.3)}$$

Downton - Found numerically due to no close form solution for $\bar{F}(t)$.

Oakes - $$\bar{H}_1(x) = \exp\left\{-\int_0^x \frac{\lambda_1 \exp(\lambda_1(\theta-1)t)}{\{\exp(\lambda_1(\theta-1)t) + \exp(\lambda_2(\theta-1)t) - 1\}} dt\right\}$$

FIGURE 1A

Relative Error in Modeling System Reliability at t_p = .25

λ_1 = 1.0, λ_2 = 1

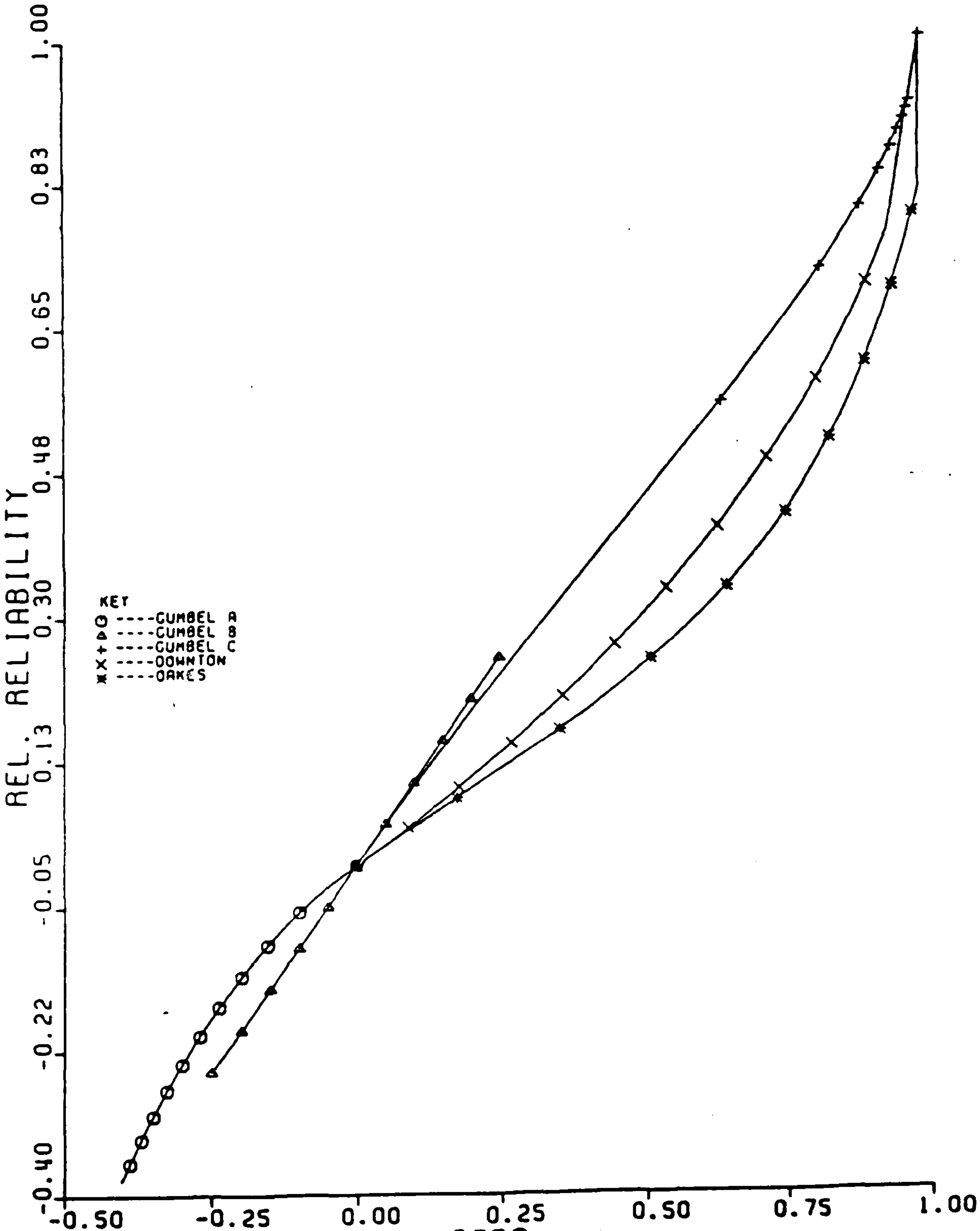


Figure 1B

Relative Error in Modeling System Reliability at t_p = .25

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

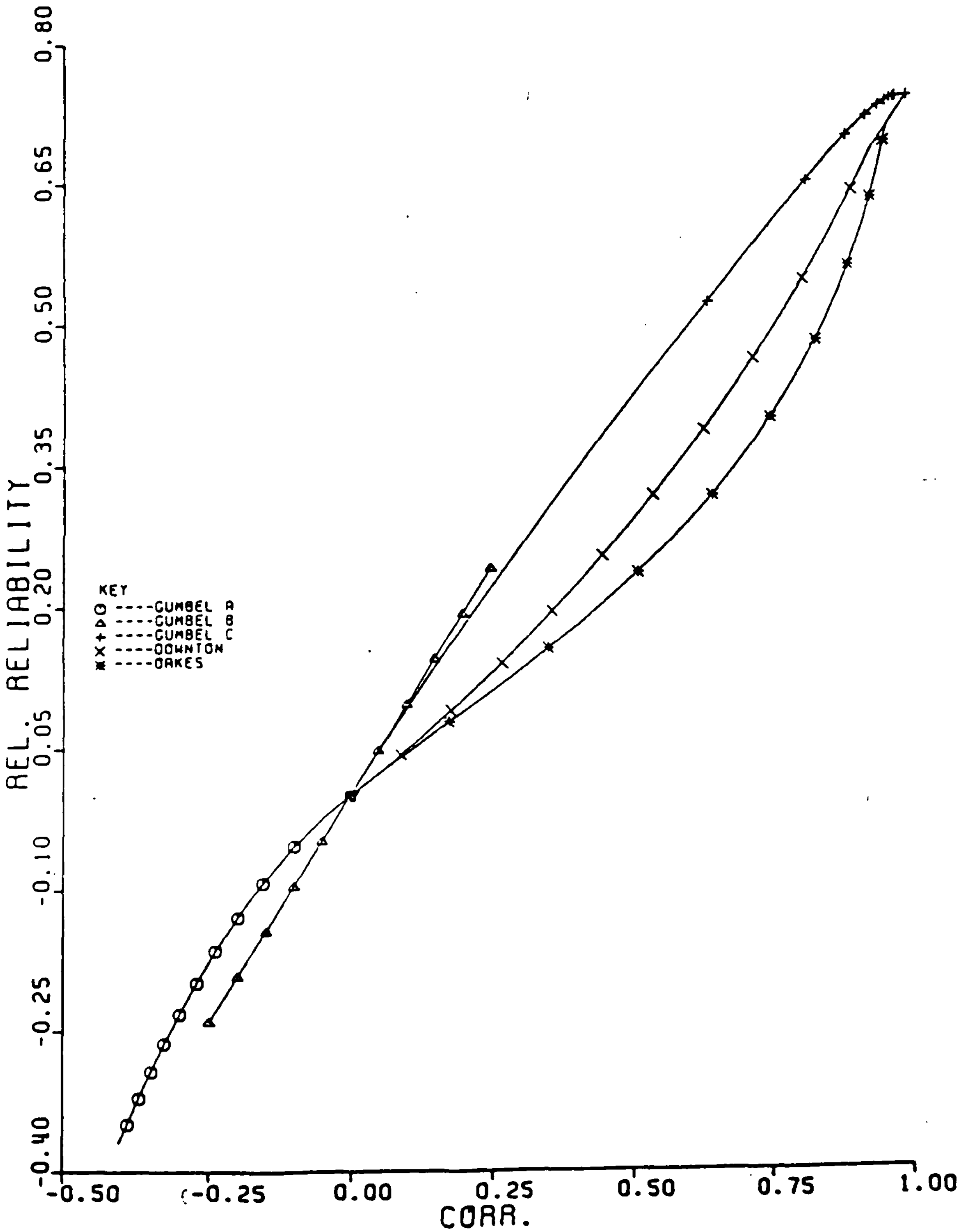


FIGURE 2A

Relative Error in Modeling System Reliability at $t_p = .5$

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

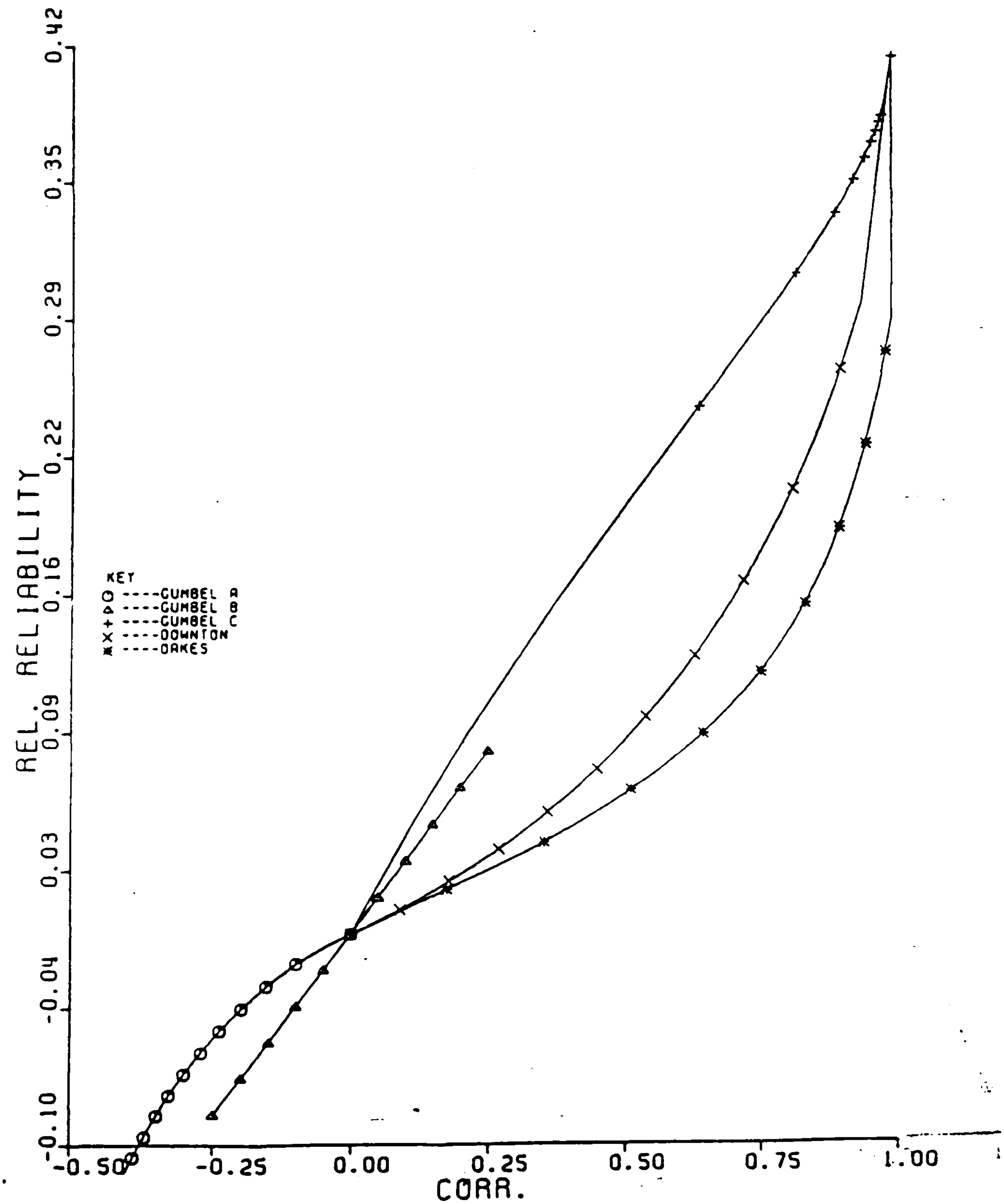


FIGURE 2B

Relative Error in Modeling System Reliability at $t_p = .5$

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

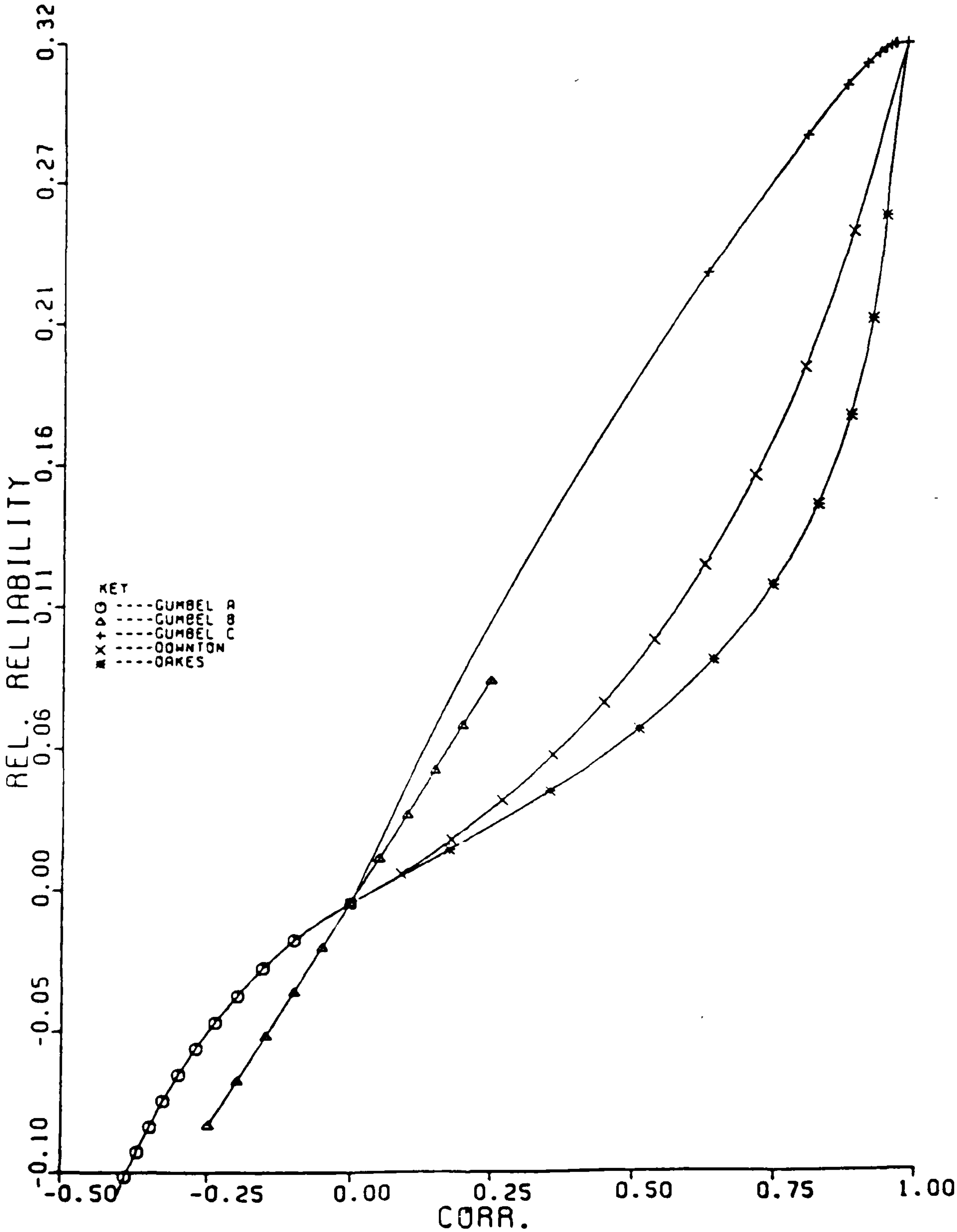


FIGURE 1C

Relative Error in Modeling System Reliability at $t_p = .75$

$\lambda_1 = 1.0,\ \lambda_2 = 1.0$

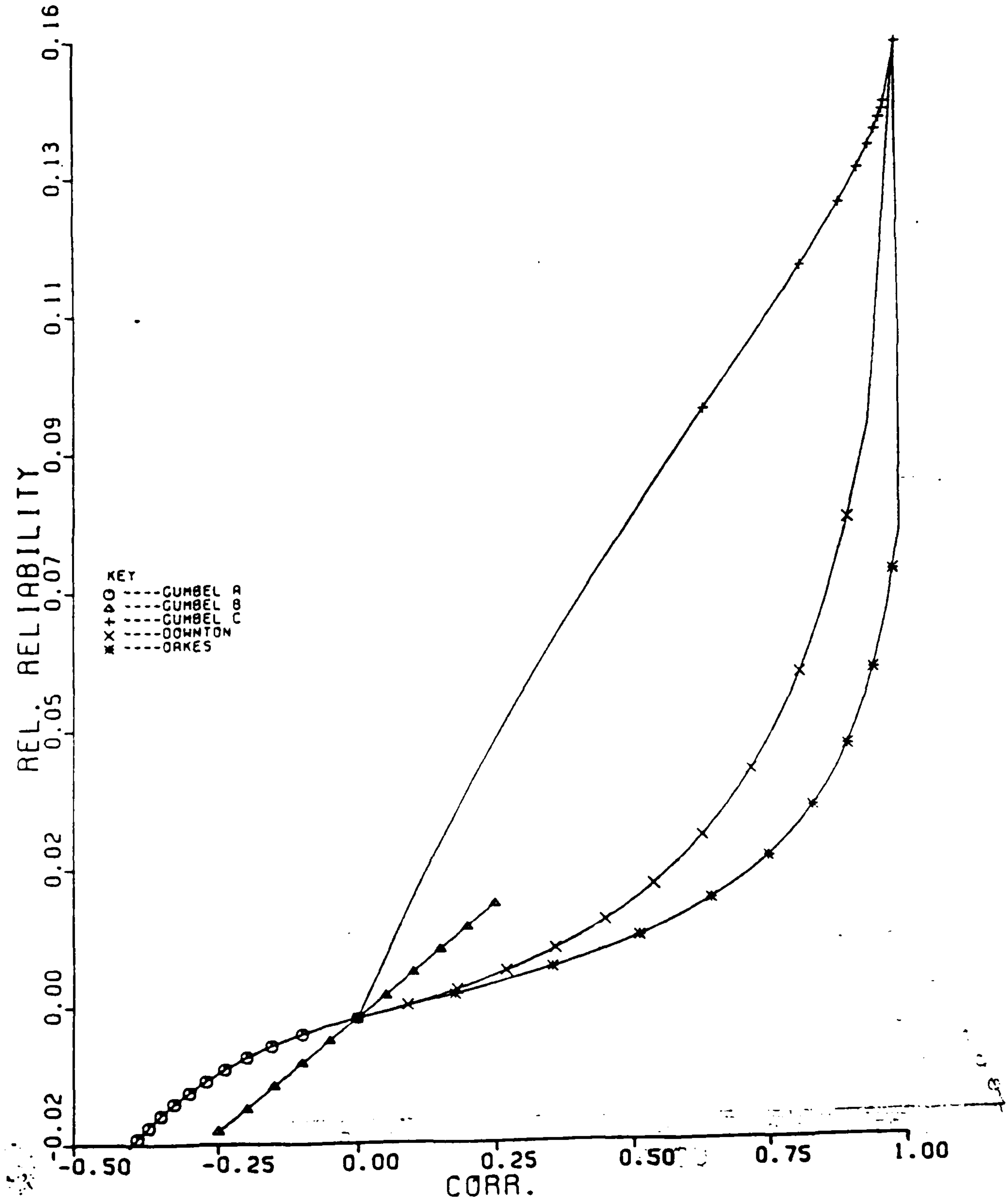


FIGURE 2C

Relative Error in Modeling System Reliability at t_p = .75

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

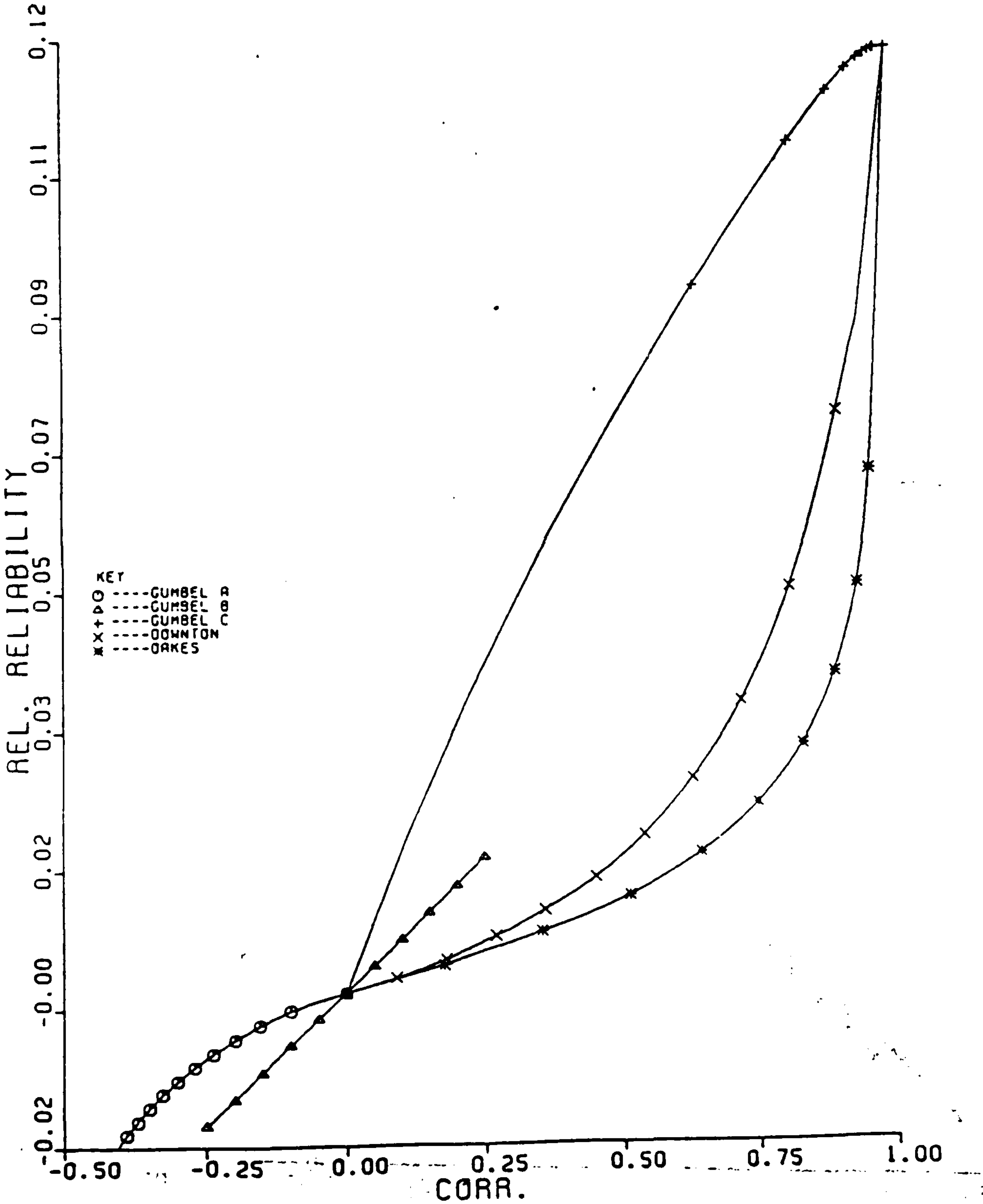


FIGURE 3A

Relative Error in Modeling Mean System Life

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

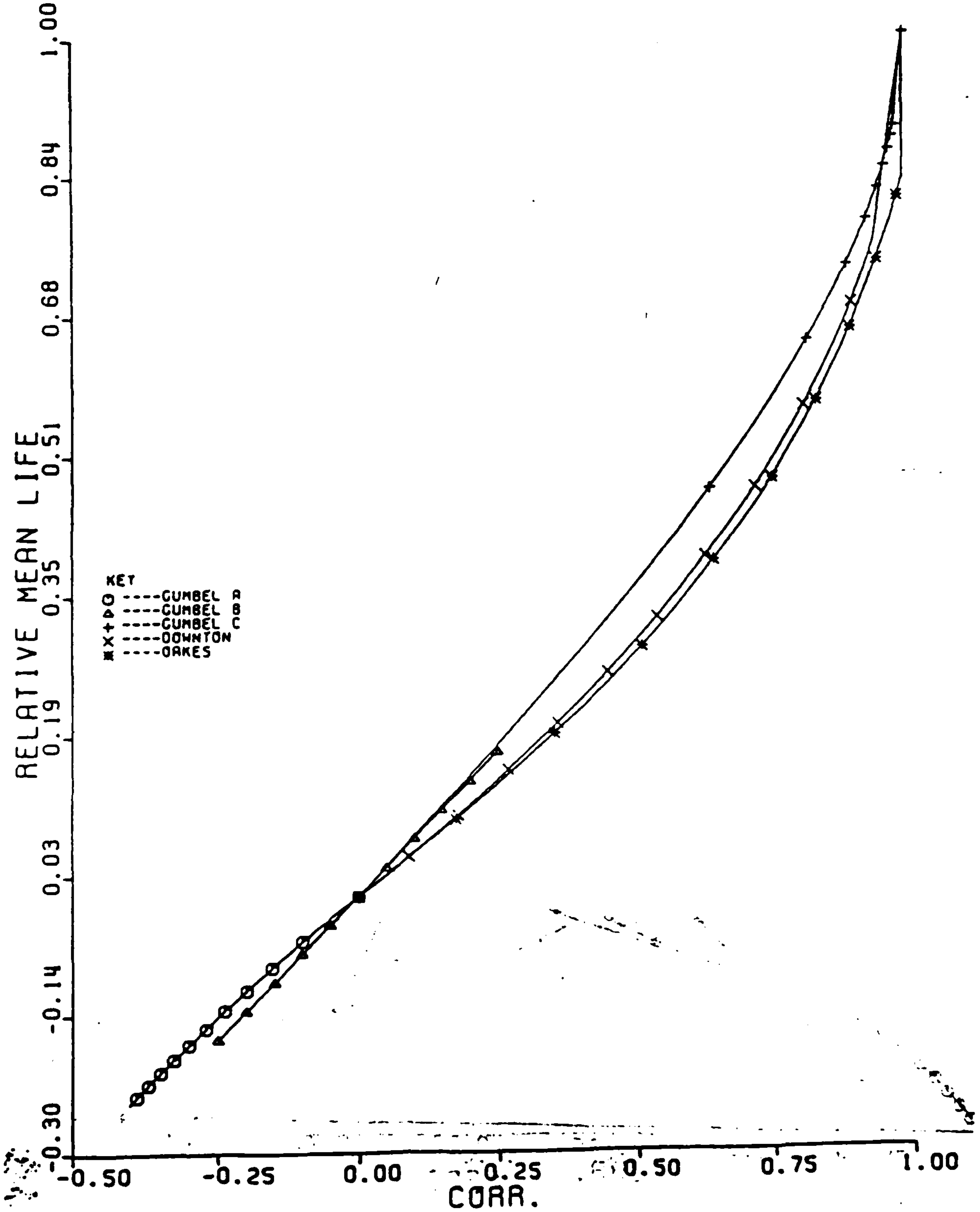


FIGURE 3B

Relative Error in Modeling Mean System Life

$\lambda_1 = 1.0,\ \lambda_2 = 1.5$

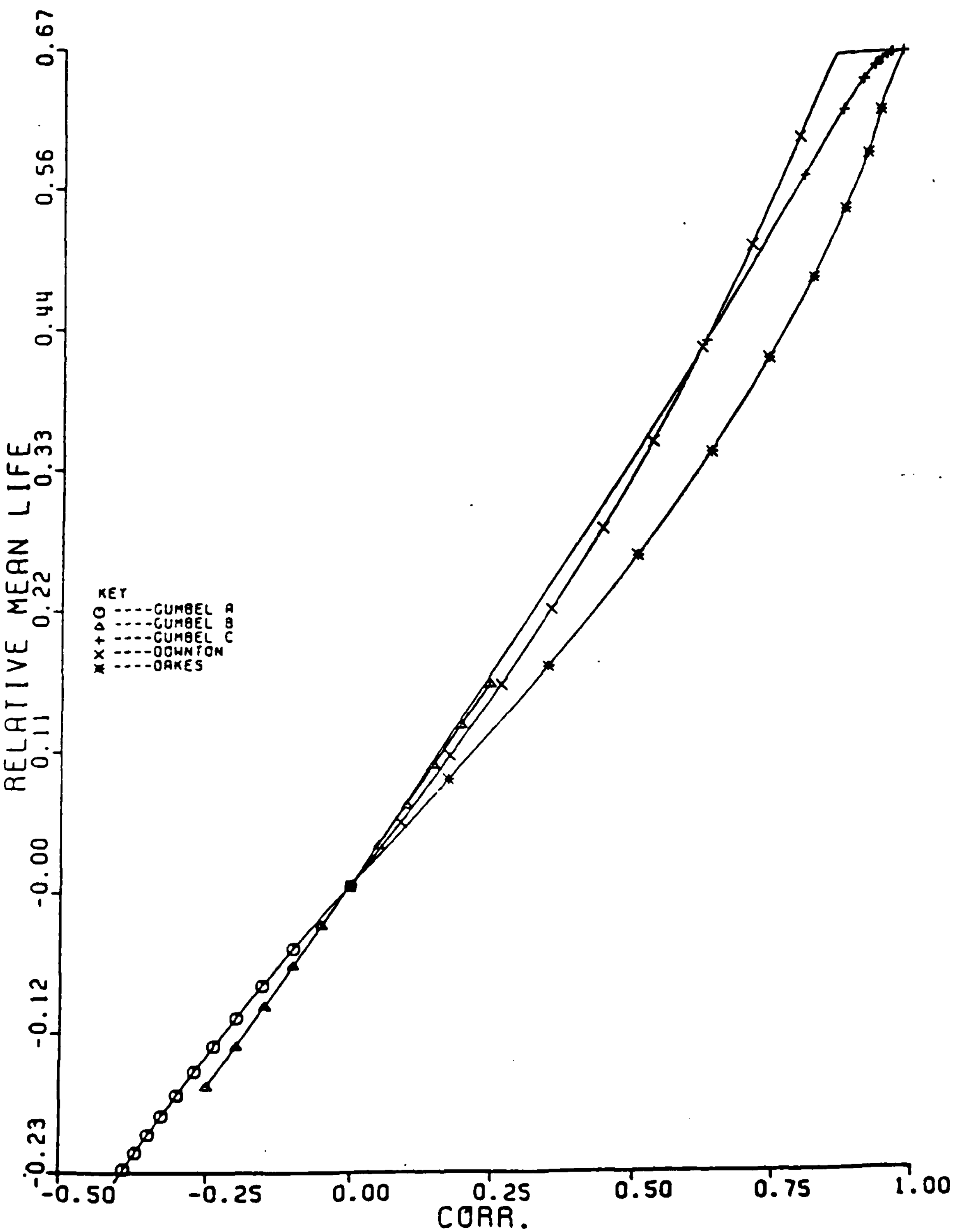


FIGURE 4A

Asymptotic Modeling Error in Estimating u_1 for K = 1.5

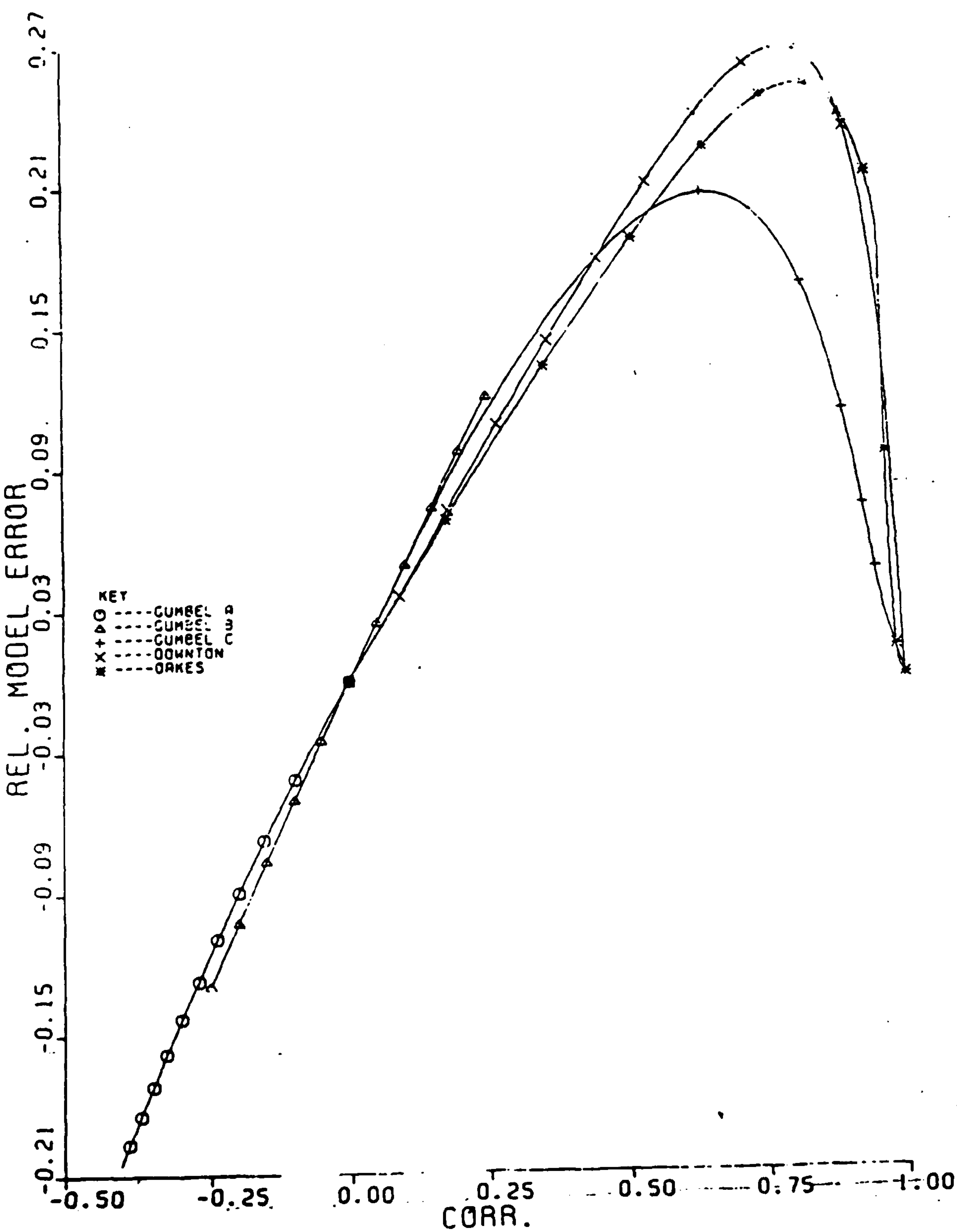


FIGURE 4B

Asymptotic Modeling Error in Estimating μ_1 for K = 1

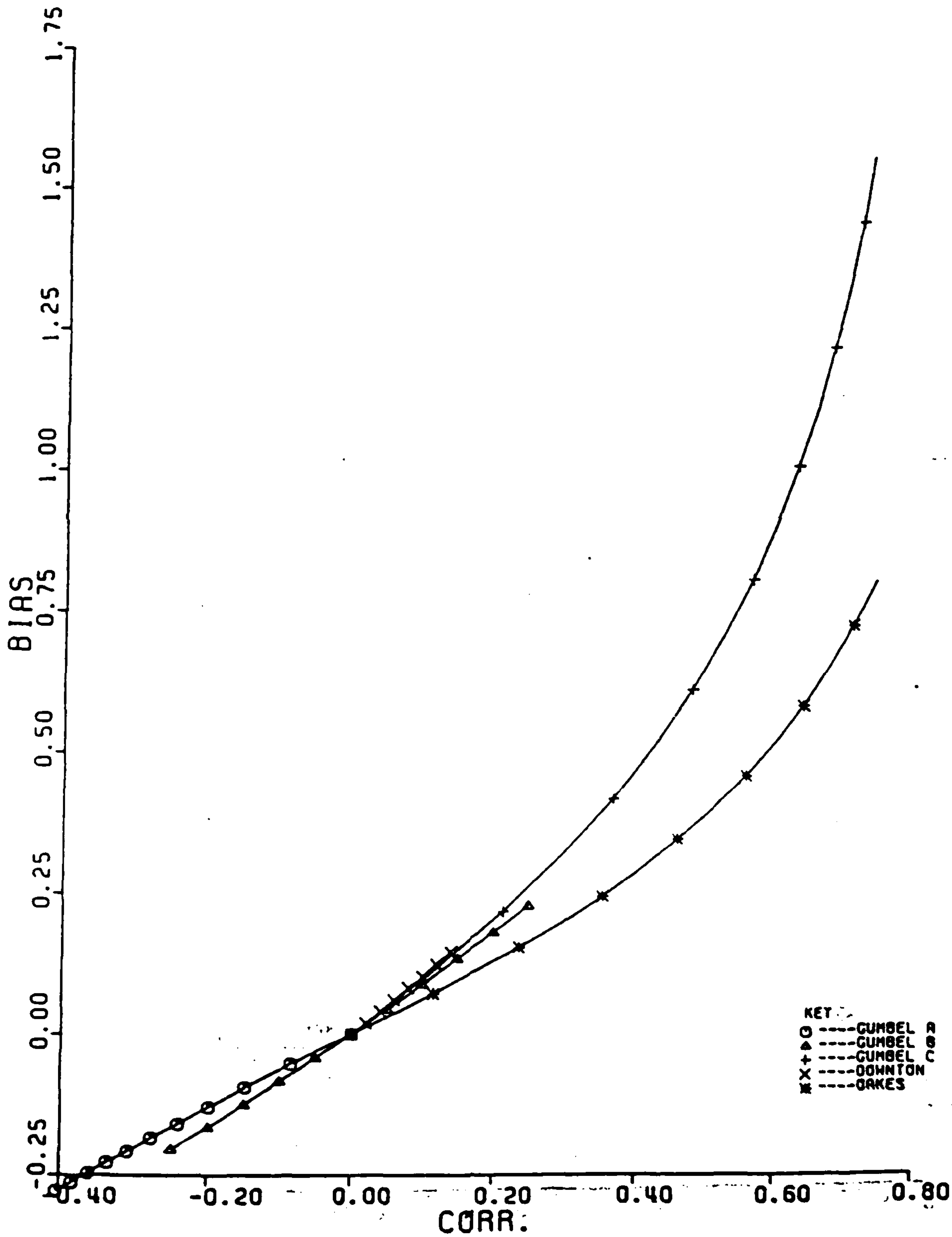


FIGURE 4C

Asymptotic Modeling Error in Estimating μ_1 for K = .67

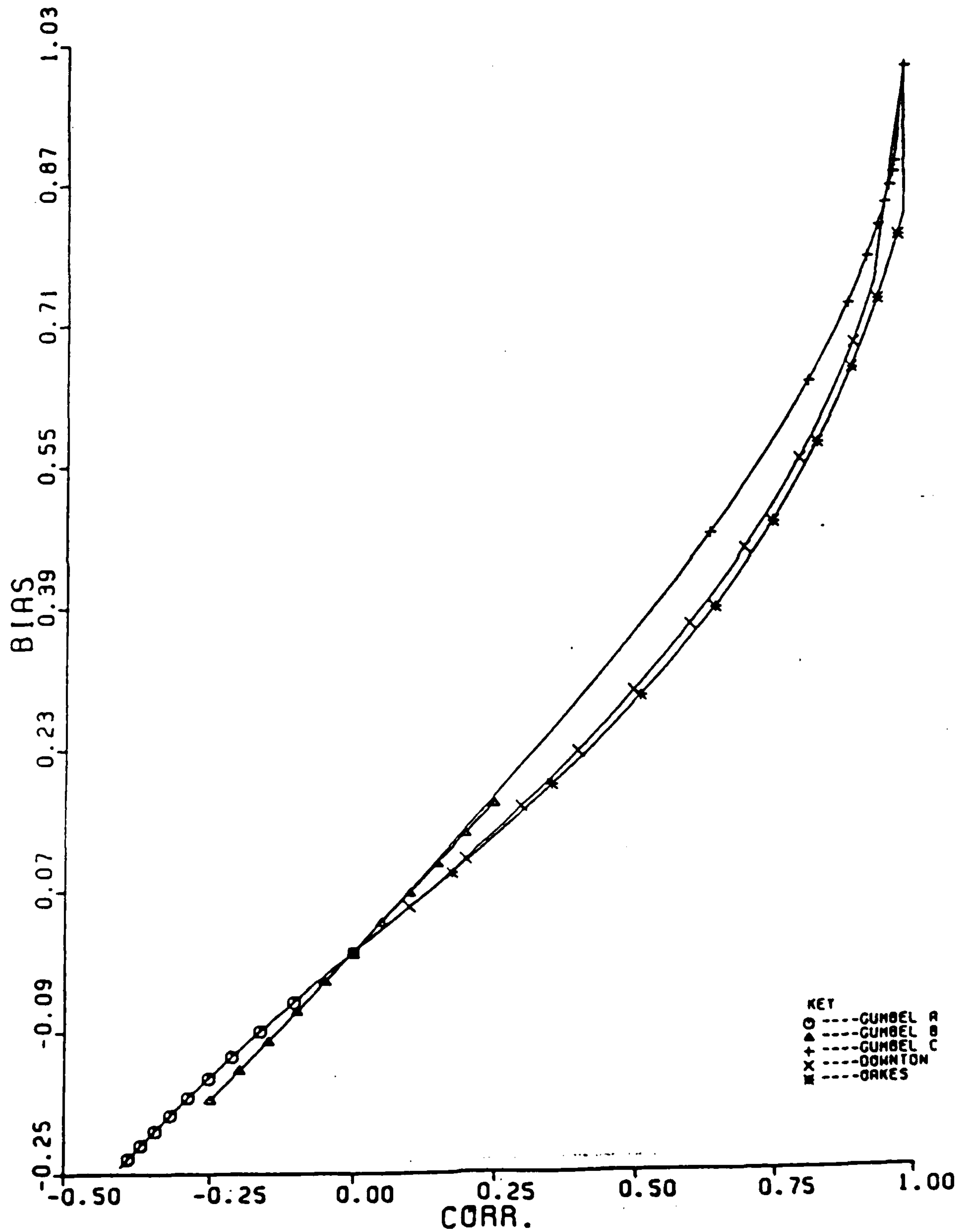


FIGURE 5A

Small Sample Size (N=10) Modeling Error in Estimating μ_1 for K = 1.5

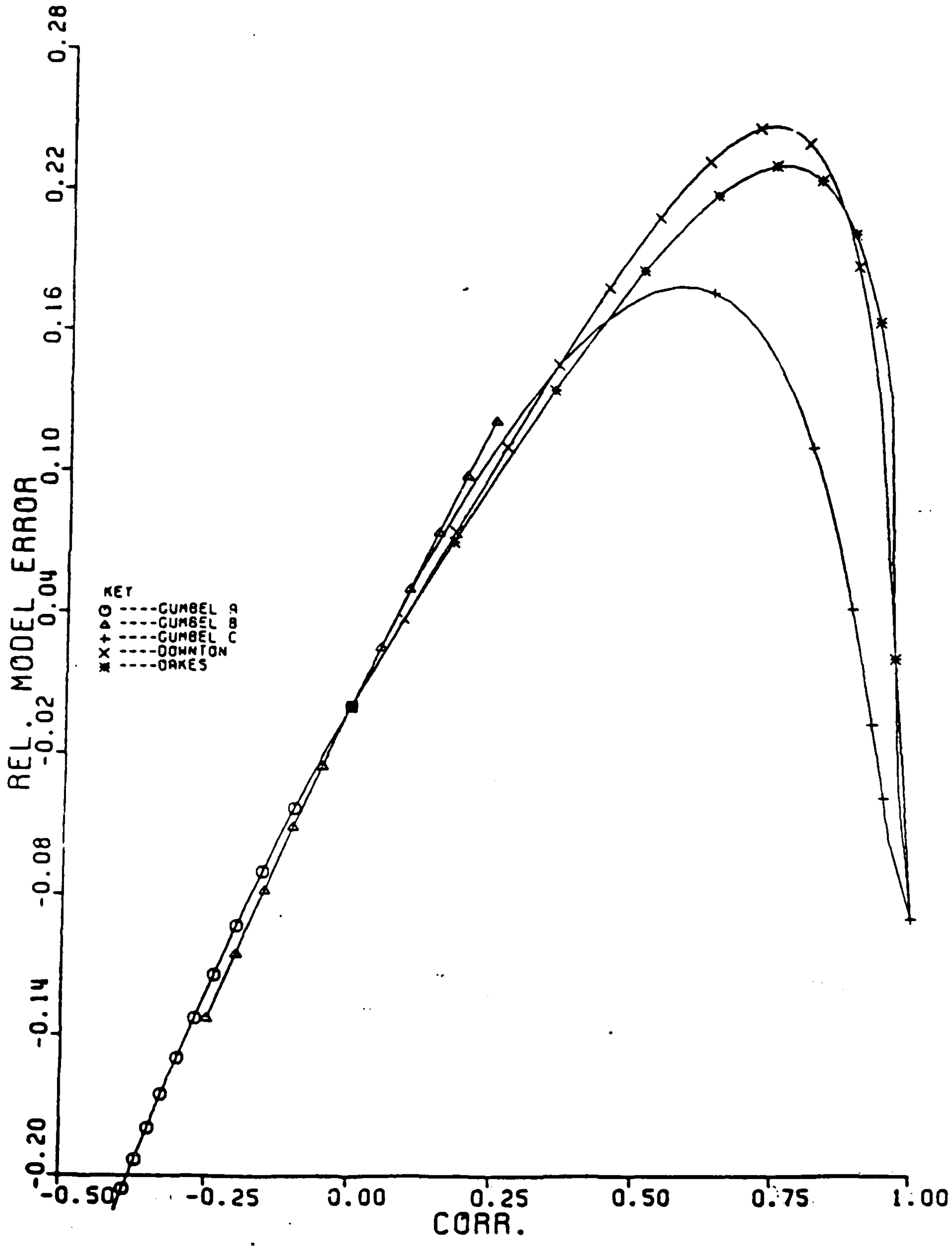


FIGURE 5B

Small Sample Size (N=10) Modeling Error in Estimating μ_1, K = 1.

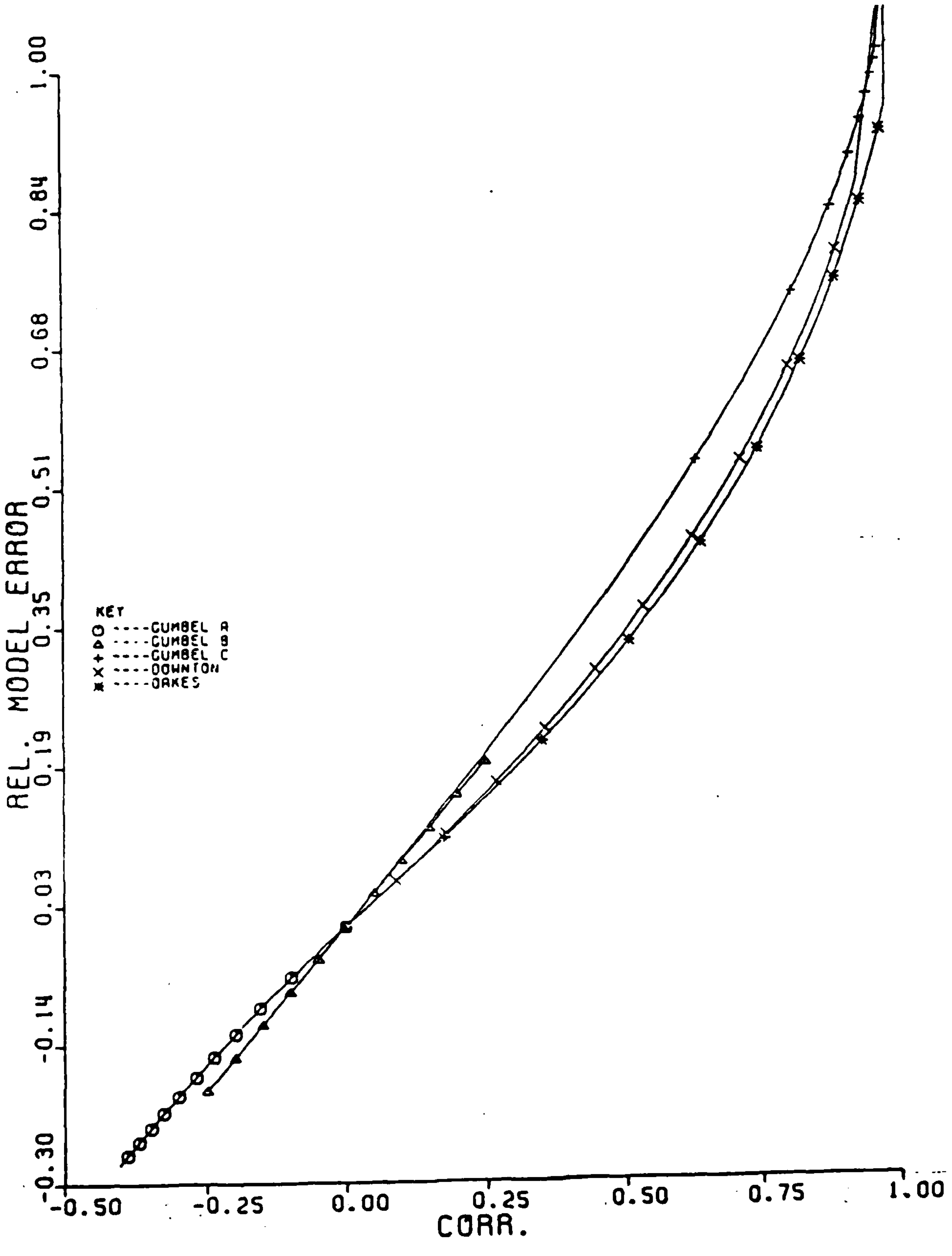


FIGURE 5C

Small Sample Size Modeling Error (N=10) in Estimating μ_1, K = .67

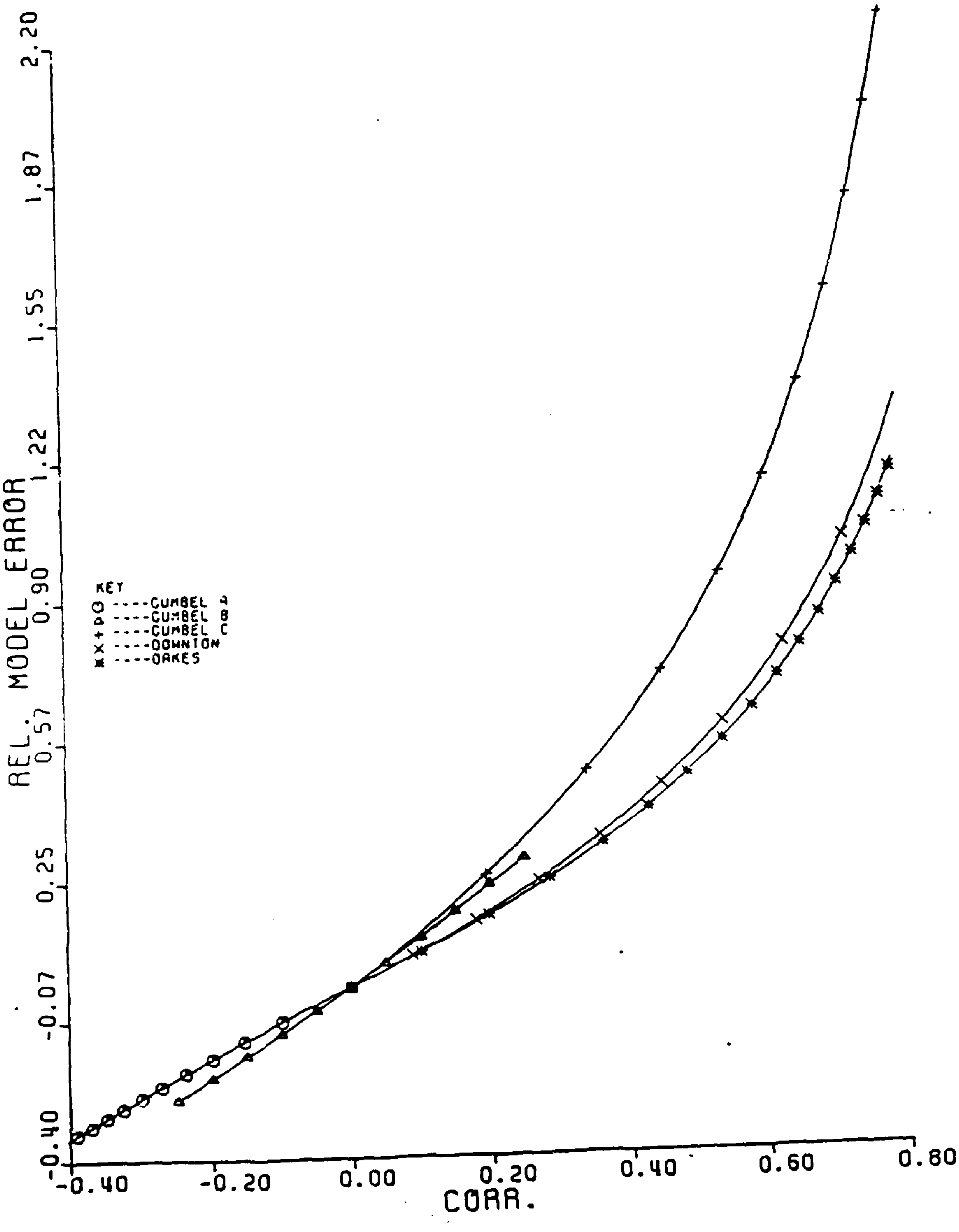


FIGURE 6A

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .25,\ K = 1.5.$

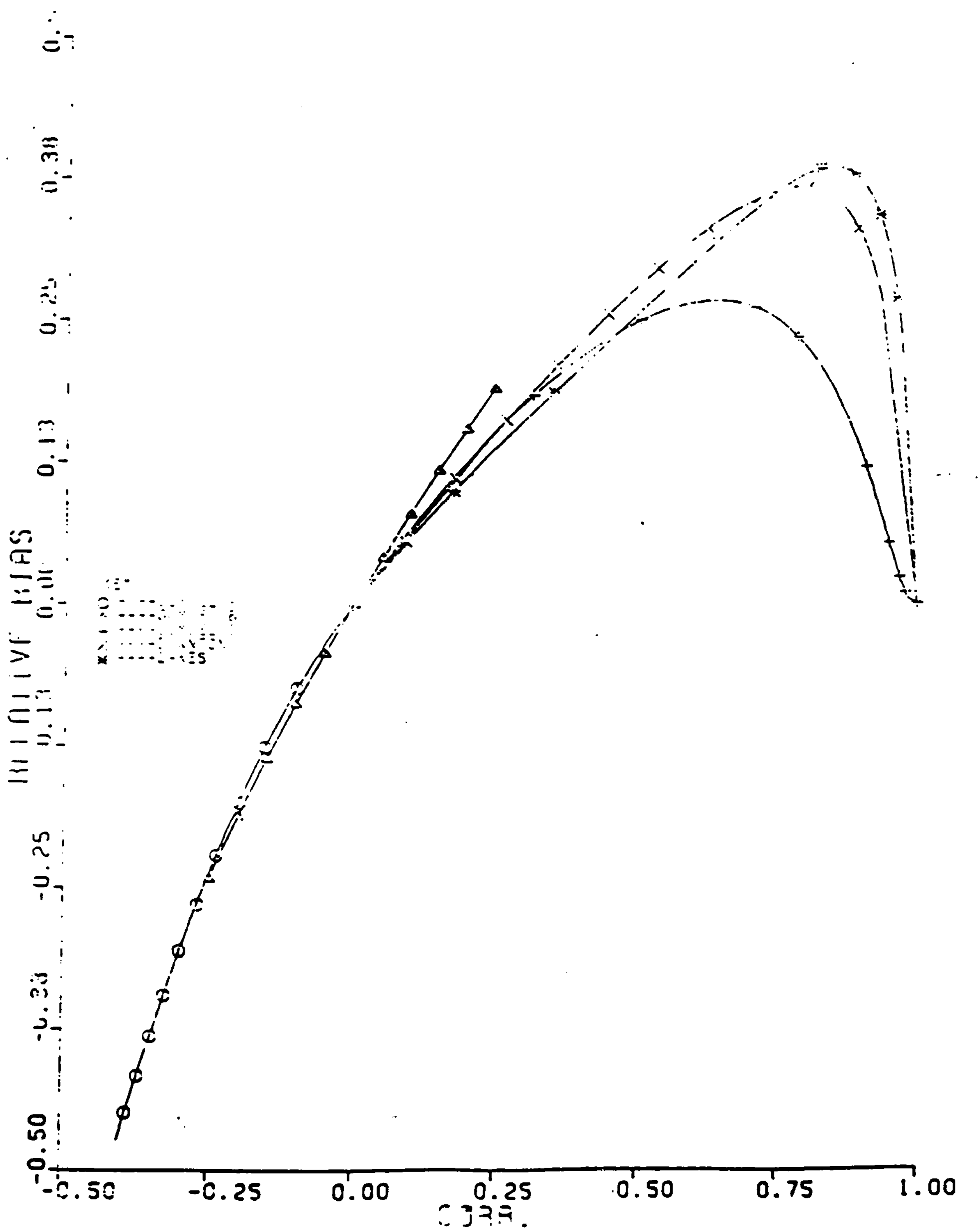


FIGURE 6B

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .25,\ K = 1.$

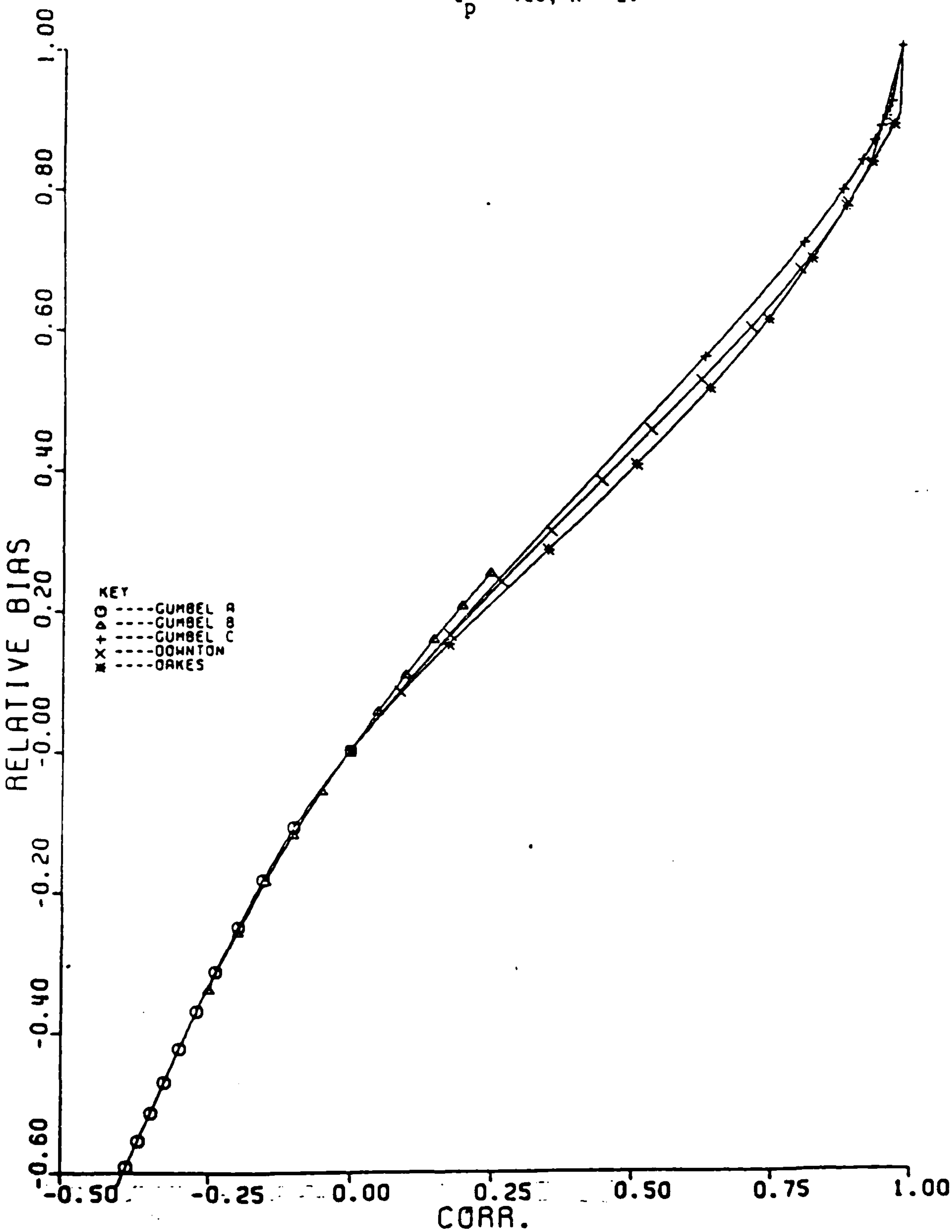


FIGURE 6C

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .25$, $K = .67$.

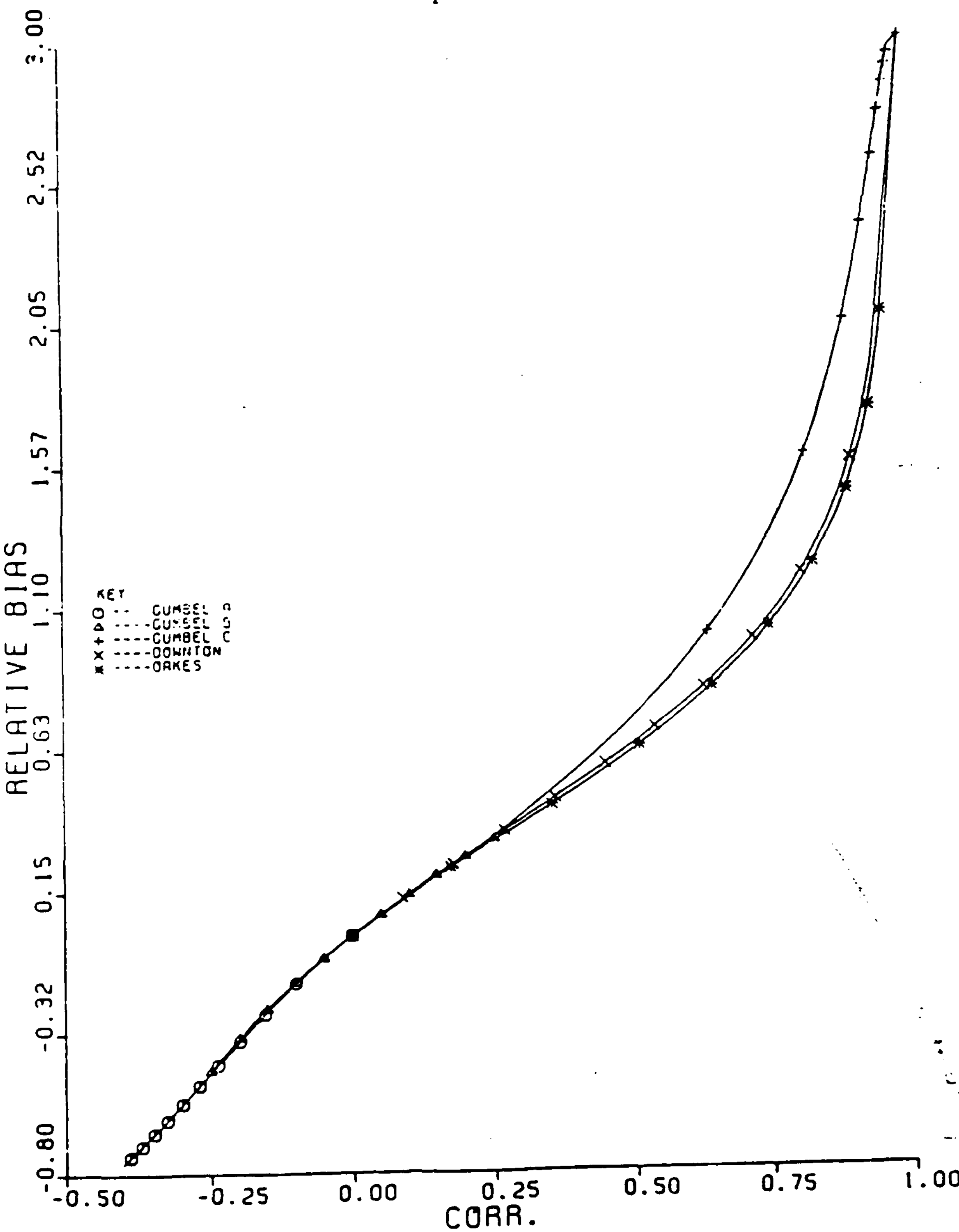


FIGURE 7A

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .25$, $K = 1.5$.

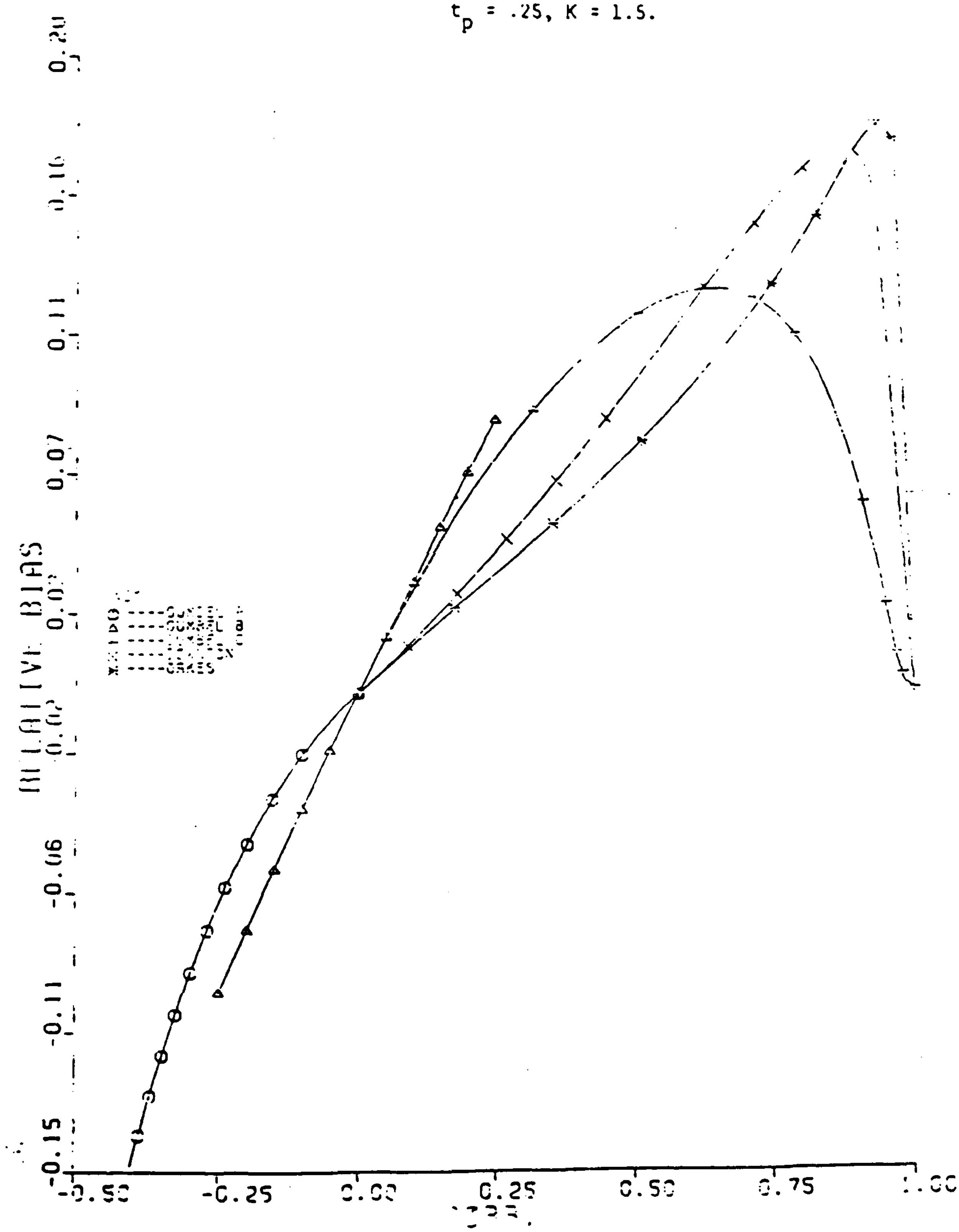


FIGURE 6B

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .5$, $K = 1$.

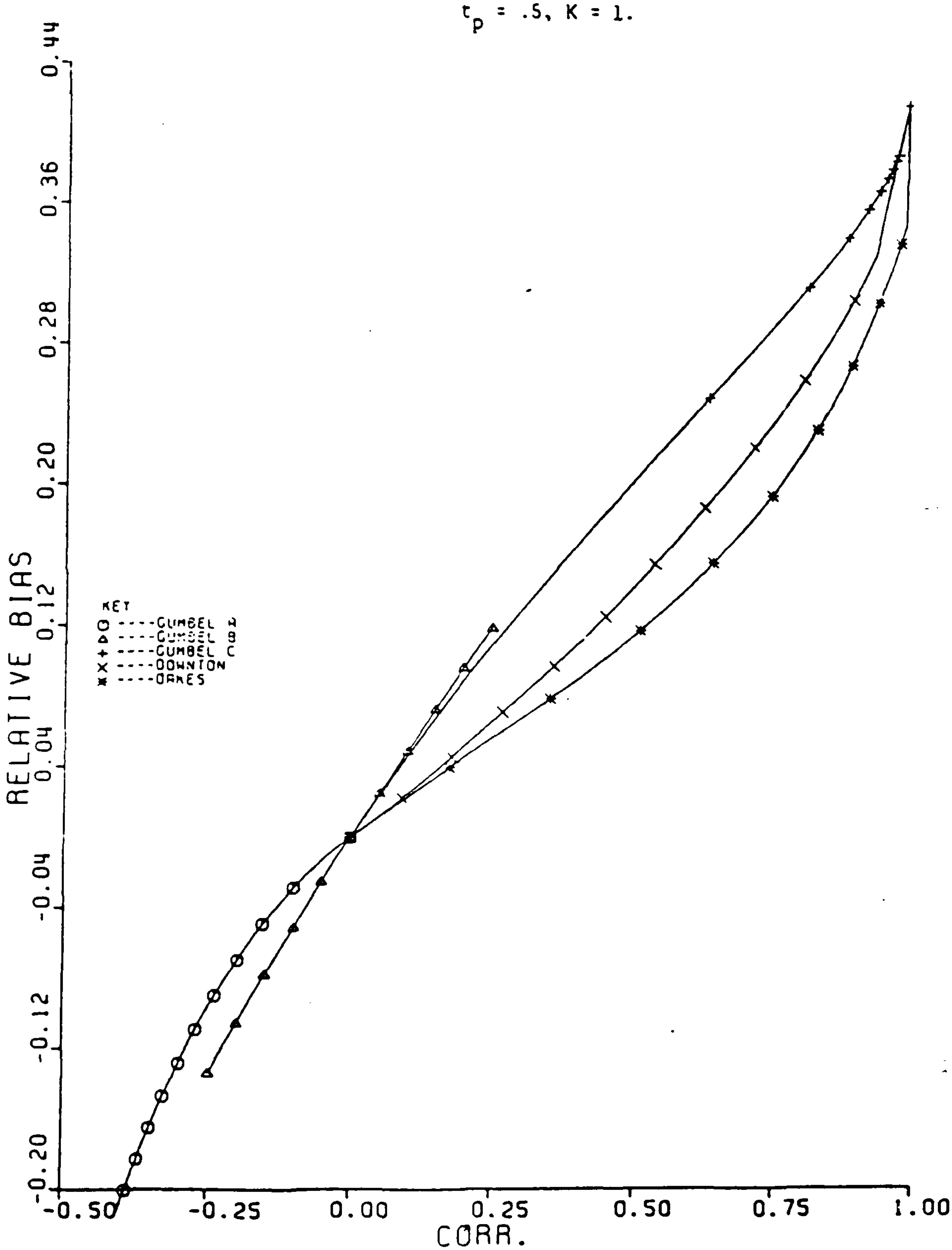


FIGURE 6C

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .5$, $K = .67$.

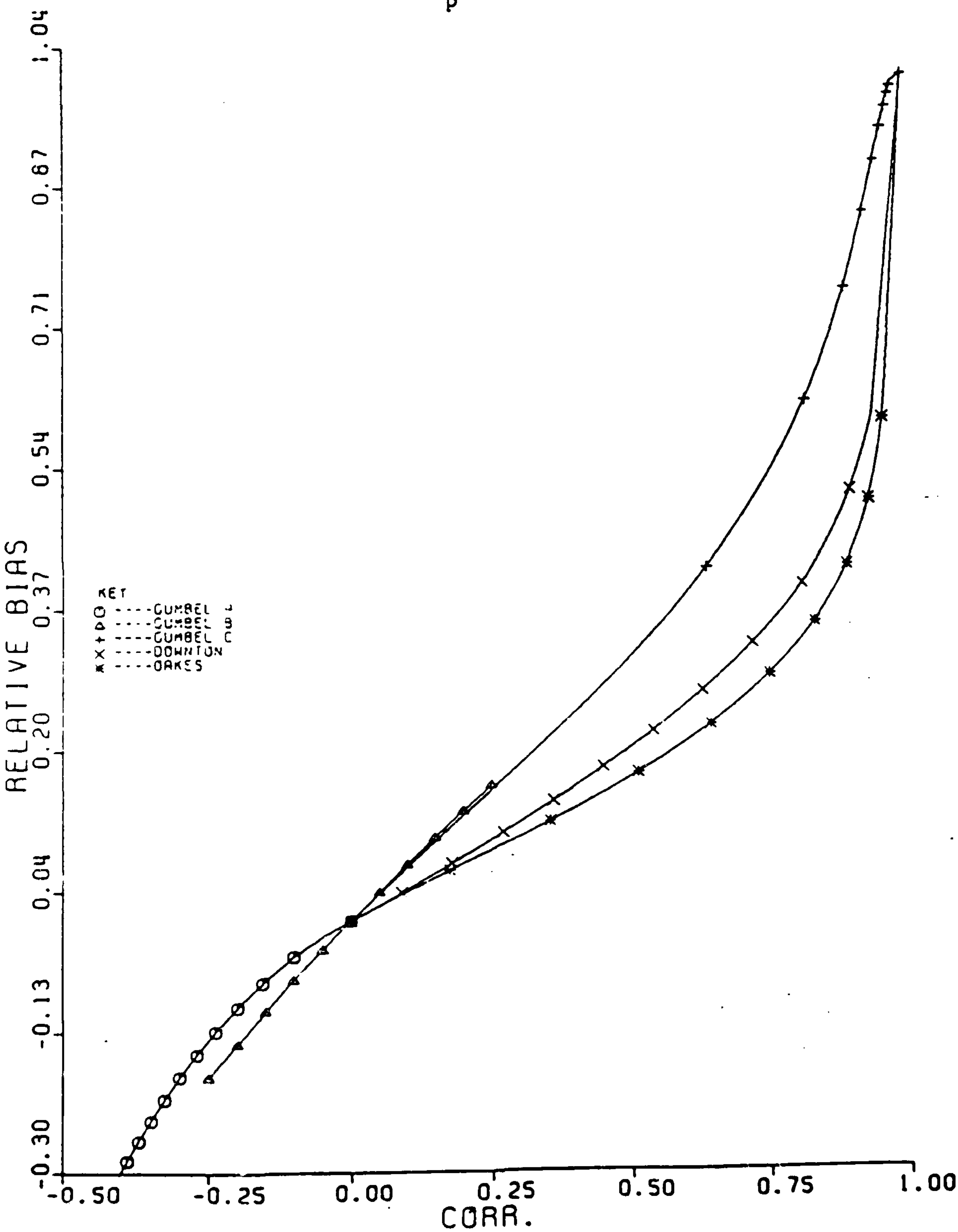


FIGURE 7A

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .75,\ K = 1.5.$

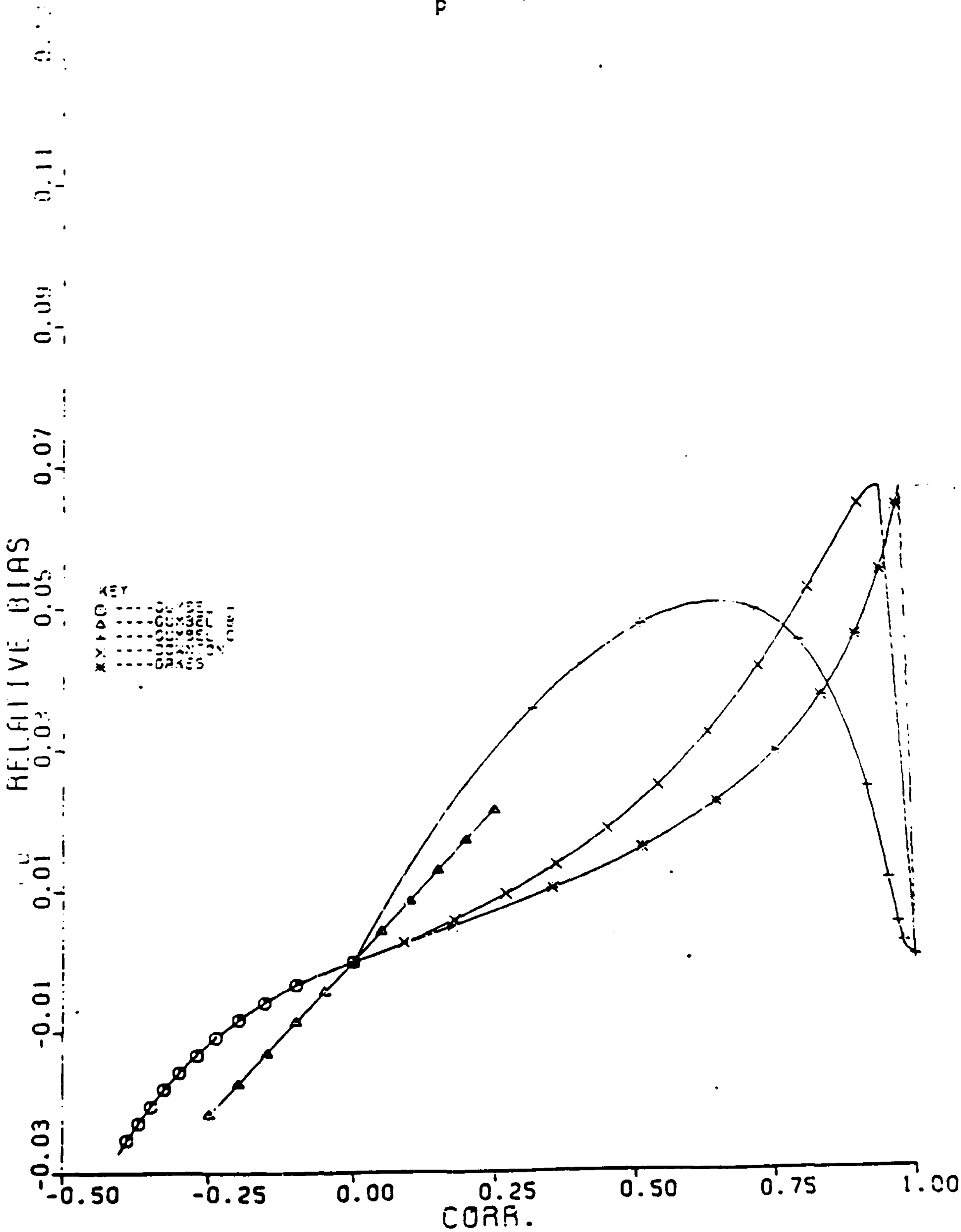


FIGURE 7B

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .75$, $K = 1$.

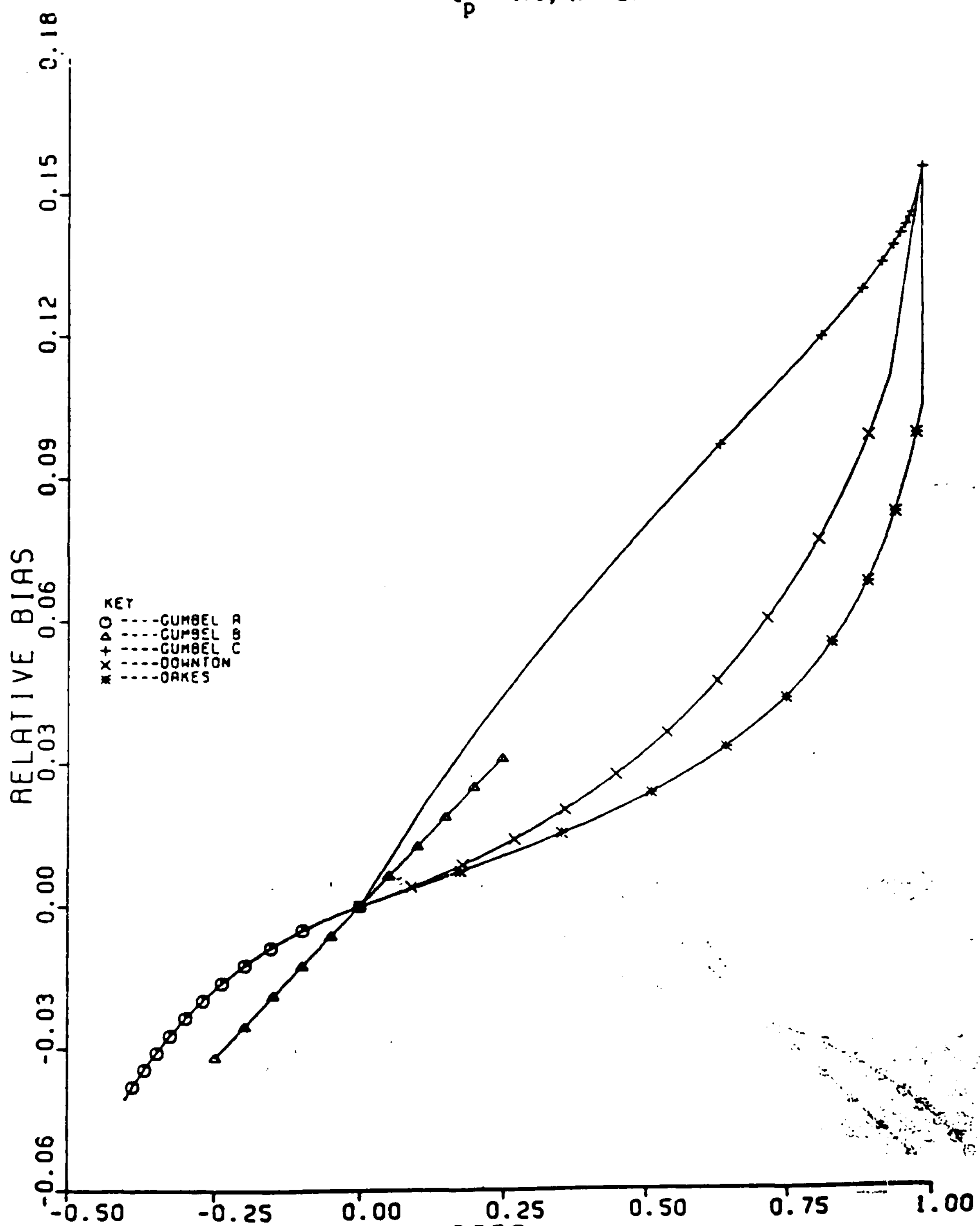
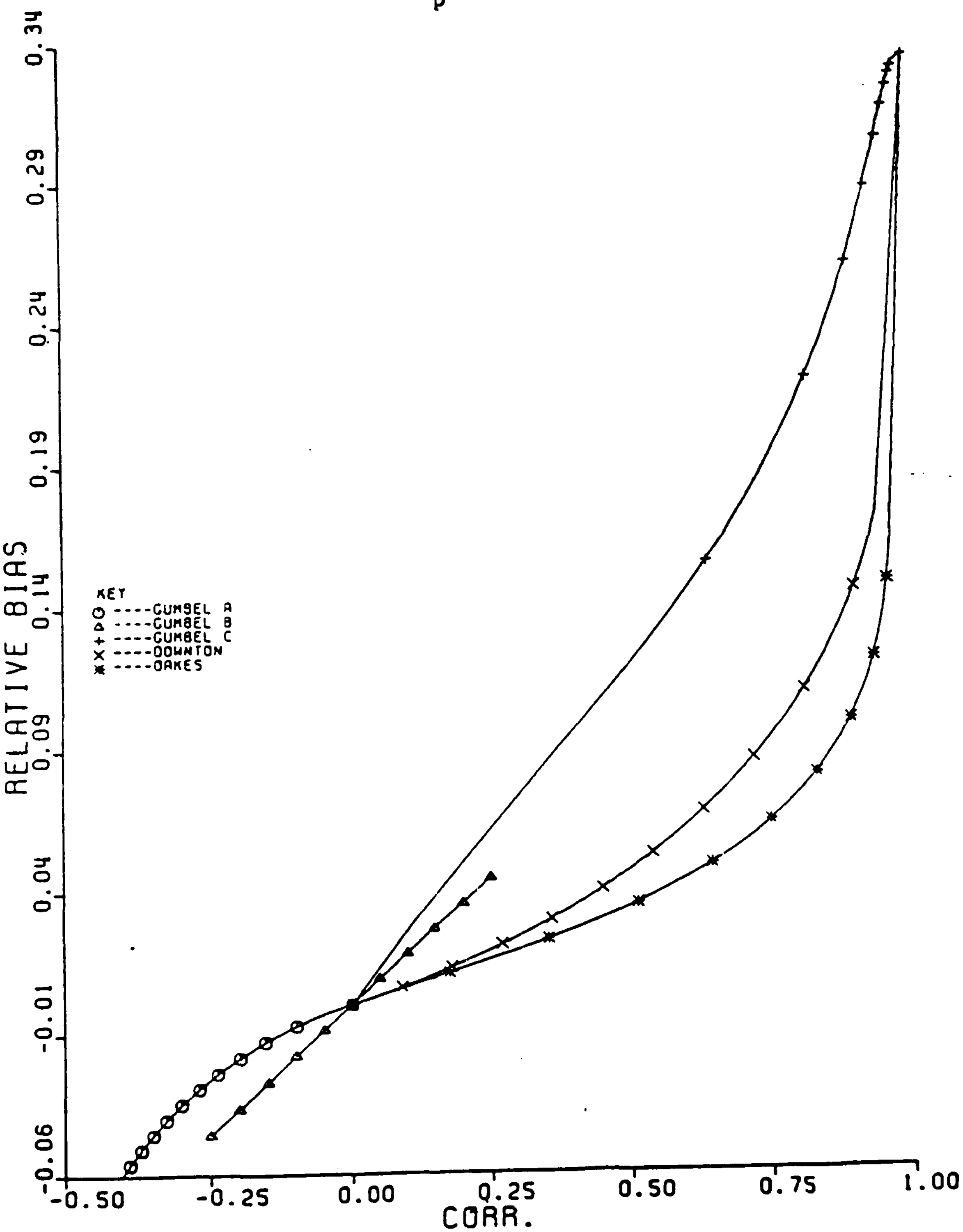


FIGURE 7C

Asymptotic Error in the Nonparametric Estimator of the First Component Survival Function at

$t_p = .75, K = .67.$



APPENDIX D

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Random Environmental Stress Model for Competing Risks		5. TYPE OF REPORT & PERIOD COVERED 10/1/84 to 10/31/85
		6. PERFORMING ORG. REPORT NUMBER 714837/763265
7. AUTHOR(s) John P. Klein and Sukhoon Lee		8. CONTRACT OR GRANT NUMBER(s) Grant No. AFOSR 82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Rd., Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 92
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington, D.C. 20332		12. REPORT DATE November 26, 1985
		13. NUMBER OF PAGES 41
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington, D.C. 20332		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A random environmental effects model is proposed for competing risks experiments. The model assumes a random stress, Z, which changes the scale parameter of each of the assumed Weibull times to occurrence of the risks. Some general properties of the model are discussed, and specific properties for a Uniform or Gamma stress model are presented. Estimation of parameters under the Gamma stress model is considered, and a new estimator based on the scaled total time on test transform is presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A RANDOM ENVIRONMENTAL STRESS MODEL FOR COMPETING RISKS

John P. Klein and Sukhoon Lee

Department of Statistics
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

A random environmental effects model is proposed for competing risks experiments. The model assumes a random stress, Z, which changes the scale parameter of each of the assumed Weibull times to occurrence of the risks. Some general properties of the model are discussed, and specific properties for a Uniform or Gamma stress model are presented. Estimation of parameters under the Gamma stress model is considered, and a new estimator based on the scaled total time on test transform is presented.

INTRODUCTION

The problem of competing risks arises naturally in a number of engineering or biological experiments. In such experiments, for some items put on test, the primary event of interests (such as death, component failure, etc.) is not observable due to the occurrence of some competing risk of removal from the study (such as censoring, failure from a different component, etc.). Competing risks arise in an engineering context in analyzing data from

(a) series systems,

(b) field tests of equipment with a fixed test time and a random or staggered entry into the study, or

(c) systems with multiple failure modes.

Competing risks arise in biological applications in analyzing data from

(a) clinical trials with a fixed trial duration and staggered entry

(b) clinical trials with some patients withdrawing from the trial prior to response

(c) studies of the time to death from a variety of causes

A common assumption made in analyzing competing risks experiments is that the potential (unobservable) times to occurrence of the competing risks are independent. This assumption is not testable due to the identifiability problem. That is, for any dependent competing risks model, there exists an independent competing risks model which yields the same observables. (See Basu and Klein (1982) for details.) However, Moeschberger and Klein (1984) show that an investigator can be appreciably misled in modeling competing risks by erroneously assuming indepencence.

In this paper we present a model for dependence between the various risks by assuming that dependence is due to some common environmental factor which effects the potential times to occurrences of each risk. In section 2 we present the model and study its properties for bivariate series and parallel systems. In section 3, we consider estimation of the model parameters for competing risks systems.

2. THE MODEL

For simplicity we shall consider the problem of bivariate systems and discuss our model in terms of engineering applications. We assume that under ideal, controlled conditions, as one may encounter in the laboratory in the testing or design stage of development, the time to failure of the two components, to be linked in a system, are X_0 and Y_0. We suppose that under these conditions, X_0, Y_0 have survival functions F_0, G_0 on $[0, \infty)$. We assume that both X_0 and Y_0 follow a Weibull form with parameters (η_1, λ_1) and (η_2, λ_2), respectively, That is, $F_0(x) = \exp(-\lambda_1 x^{\eta_1})$. The Weibull distribution, which may have increasing ($\eta > 1$), decreasing ($\eta < 1$) or constant failure rate ($\eta = 1$) has been shown experimentally to provide a reasonable fit to many different types of survival data. (See Bain (1978)). We now link the two components into a system in such a way that under ideal lab conditions the two components are independent.

Now suppose that the above system (X_0, Y_0) is put into operation under usage conditions. We suppose that under such conditions the effect of the environment is to degrade or improve each component by the same random amount. That is, the effect of the environment is to select a random factor, Z, from some distribution, H, which changes the maginal survival functions of the two components to F_0^Z and G_0^Z. A value of Z less than one means that component reliabilities are simultaneously improved, while a value of Z greater than one implies a joint degradation. The resulting joint reliability of the two components' lifetimes, (X,Y) in the operating environment is

$$F(x,y) = E[\exp(-Z(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}))]. \tag{2.1}$$

This model has been proposed by Lindley and Singpurwalla (1984) in the reliability context when F_0, G_0 are exponential and H( ) follows a gamma distribution. This basic dependence structure was also proposed by Clayton (1978) to model associations in bivariate survival data, and later by Oakes (1982) to model bivariate survival data. Hutchinson (1982) proposed a similar model when H( ) has a gamma distribution and $F_0(t) = G_0(t) = \exp(-t^{\eta})$.

The model described above for a general distribution of the environmental stress has a particular dependence structure which we summarize in the following lemmas.

Lemma 1. Let (X,Y) follow the model (2.1) where Z is a positive random variable with finite $\left(\frac{r}{\eta_1} + \frac{s}{\eta_2}\right)^{th}$ inverse moment. Then

$$E(X^r Y^s) = \lambda_1^{-r/\eta_1} \lambda_2^{-s/\eta_2} \Gamma(1 + r/\eta_1)\ \Gamma(1 + s/\eta_2)\ E(Z^{-(r/\eta_1 + s/\eta_2)}) \tag{2.2}$$

The proof follows by noting that, given $Z = z$, (X,Y) are independent Weibulls with parameters $(\eta_1, \lambda_1 z)$ and $(\eta_2, \lambda_2 z)$, respectively and $E(X^r|Z=z) = \lambda_1^{-r/\eta_1} z^{-r/\eta_1} \Gamma(1+ r/\eta_1)$ with a similar expression for Y^s. When the appropriate moments exist, we have

(A) $E(X) = E(X_0)\, E(Z^{-1/\eta_1})$,

(B) $V(X) = E(X_0^2)\, \mathrm{Var}(Z^{-1/\eta_1}) + E(Z^{-1/\eta_1})^2\, \mathrm{Var}(X_0)$,

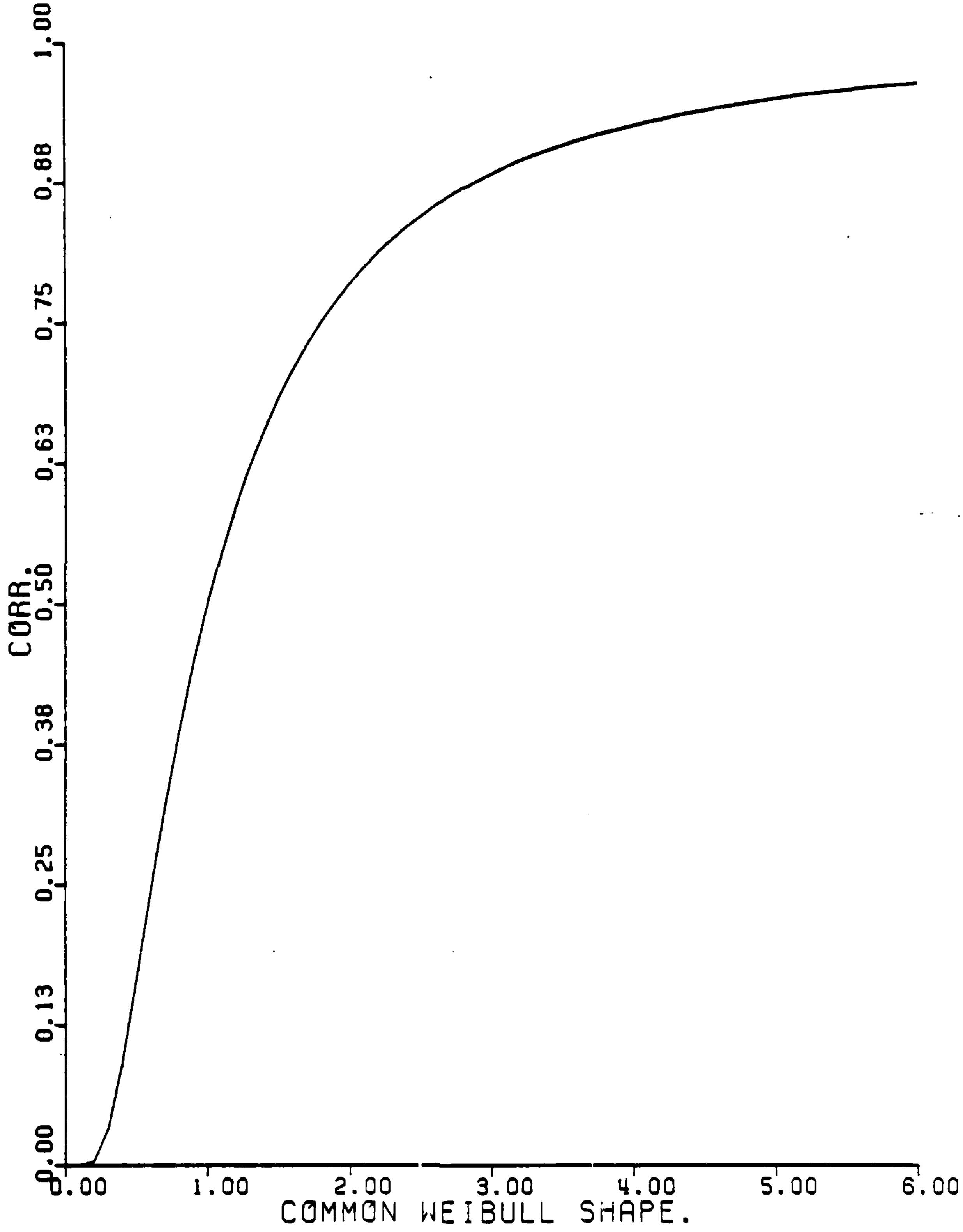
(C) $\mathrm{Cov}(X,Y) = E(X_0)\, E(Y_0)\, \mathrm{Cov}(Z^{-1/\eta_1}, Z^{-1/\eta_2})$ which is greater than 0.

If $\eta_1 = \eta_2 = \eta$ then the correlation between (X,Y) is

$$\rho = \frac{\Gamma(1+ 1/\eta)^2\, \mathrm{Var}(Z^{-1/\eta})}{\mathrm{Var}(Z^{-1/\eta})\, \Gamma(1+2/\eta) + (\Gamma(1+2/\eta) - \Gamma(1+1/\eta)^2)\, E(Z^{-1\eta})^2}.$$

In this case the correlation is bounded above by $\Gamma(1+1/\eta)^2 / \Gamma(1+2/\eta)$. Figure 1 shows the maximal correlation as a functcion of η for $\eta \in (0, 10)$. Note that this maximal correlation is an

FIGURE 1
UPPER BOUND ON MAXIMAL CORRELATION FOR RANDOM ENVIORNMENT MODEL.



increasing function of η. One can also show that F(x, y) is positive quadrant dependent for any η_1, η_2.

Exact expressions for competing risks quantities of interest can be computed when a particular model is assumed for the distribution of Z. We shall consider the gamma and uniform models. Consider first the gamma model with $h_z(z) = b^a z^{a-1} \exp(-bz)/\Gamma(a)$, $z > 0$. For this model, the joint survival function is

$$F(x,y) = \frac{b^a}{[b+\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}]^a} \tag{2.3}$$

which is a bivariate Burr Distribution (see Takahasi (1965)), the marginal distributions are univariate Burr distributions with

$$E(X) = (\lambda_1/b)^{-1/\eta_1} \Gamma(1+1/\eta_1)\Gamma(a-1/\eta_1)\,\Gamma(a), \text{ if } a > 1/\eta,$$

$$Var(X) = (\lambda_1/b)^{-2/\eta_1} \{ \frac{\Gamma(1+2/\eta_1)\Gamma(a-2/\eta_1)}{\Gamma(a)} - [\frac{\Gamma(1+1/\eta_1)\Gamma(a-1/\eta_1)}{\Gamma(a)}]^2\}, \text{ if } a > 2/\eta_1$$

with similar expressions for E(Y), Var(Y). The covariance of (X,Y) is

$$Cov(X,Y) = (\lambda_1/b)^{-1/\eta_1} (\lambda_2/b)^{-1/\eta_2} \Gamma(1+1/\eta_1)\,\Gamma(1+1/\eta_2)\{\frac{\Gamma(a-1/\eta_1-1/\eta_2)}{\Gamma(a)} - \frac{\Gamma(a-1/\eta_2)\,\Gamma(a-1/\eta_2)}{\Gamma(a)}\}$$

for $a > 1/\eta_1 + 1/\eta_2$. For the gamma model, the reliability function for a bivariate series system is given by

$$R_s(t) = (1+(\lambda_1/b)t^{\eta_1} + (\lambda_2/b)t^{\eta_2})^{-a}, \tag{2.4}$$

and for a parallel system by

$$R_p(t) = (1+(\lambda_1/b)t^{\eta_1} + (1+(\lambda_2/b)t^{\eta_2})^{-a} - (1+(\lambda_1/b)t^{\eta_2} + (\lambda_2/b)t^{\eta_2})^{-a} \tag{2.5}$$

Figures 2A-E are plots of the series system reliability for $\lambda_1 = 1$, $\lambda_2 = 2$ and several combinations of η_1, η_2. Each figure shows the reliablity for a = 1/2, 1, 2, 4, and the independent Weibull model. In all cases, b = 1. For these figures we note that for fixed $\lambda_1, \lambda_2\ \eta_1, \eta_2$, t, the series system reliability is a decreasing function of the shape parameter a. Figures 3A-E are plots of the parallel system reliability (2.5) for the above parameters. Again, the reliability is a decreasing function of a. Also in both the series and parallel system reliability, the shape of the reliability function is quite different from that encountered under independence.

The gamma model is a reasonable model for the environmental stress due to its flexibility and the tractability of the model in obtaining close form solutions for the relevant quantities and in estimating parameters. However, in some cases, such as when the operating environment is always more severe than the laboratory environment, the support of H may be restricted to some fixed interval. A possible model for such an environmental stress is the uniform distribution over [a,b]. For this model, the joint survival function is

$$F(x,y) = \frac{[\exp(-b(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2})) - \exp(-a(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2}))]}{(b-a)(\lambda_1 x^{\eta_1} + \lambda_2 y^{\eta_2})} \tag{2.6}$$

$$E(X) = \lambda_1^{-1/\eta_1}\,\Gamma(1+1/\eta_1)\,\eta_1\,(b^{(\eta_1-1)/\eta_1} - a^{(\eta_1-1)/\eta_1})/\{(\eta_1-1)(b-a)] \qquad \text{if } \eta_1 \neq 1$$

$$= \ln(b/a)/[\lambda_1(b-a)] \qquad \text{if } \eta_1 = 1,$$

$$Var(X) = \frac{\eta_1\,\lambda_1^{-2/\eta_1}}{(b-a)}\{\,\Gamma(1+2/\eta_1)\,\eta_1\,(b^{(\eta_1-2)/\eta_1} - a^{(\eta_1-2)/\eta_1})$$

$$\frac{-\,\Gamma(1+1/\eta_1)^2\,\eta_1\,(b^{(\eta_1-1/\eta_1)} - a^{(\eta_1-1/\eta_1)})^2}{(\eta_1-1)^2(b-a)}\} \qquad \text{if } \eta_1 \neq 1,2$$

$$2/(\lambda_1^2 ab) - \ln(b/a)^2/[(b-a)\,\lambda_1]^2 \qquad \text{if } \eta_1 = 1$$

$$\lambda_1^{-1}\left[\frac{\ln(b/a)}{(b-a)} - \frac{\Pi}{(b^{1/2}+a^{1/2})^2}\right] \qquad \text{if } \eta_1 = 2$$

FIGURE 2 A

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=1.0.$

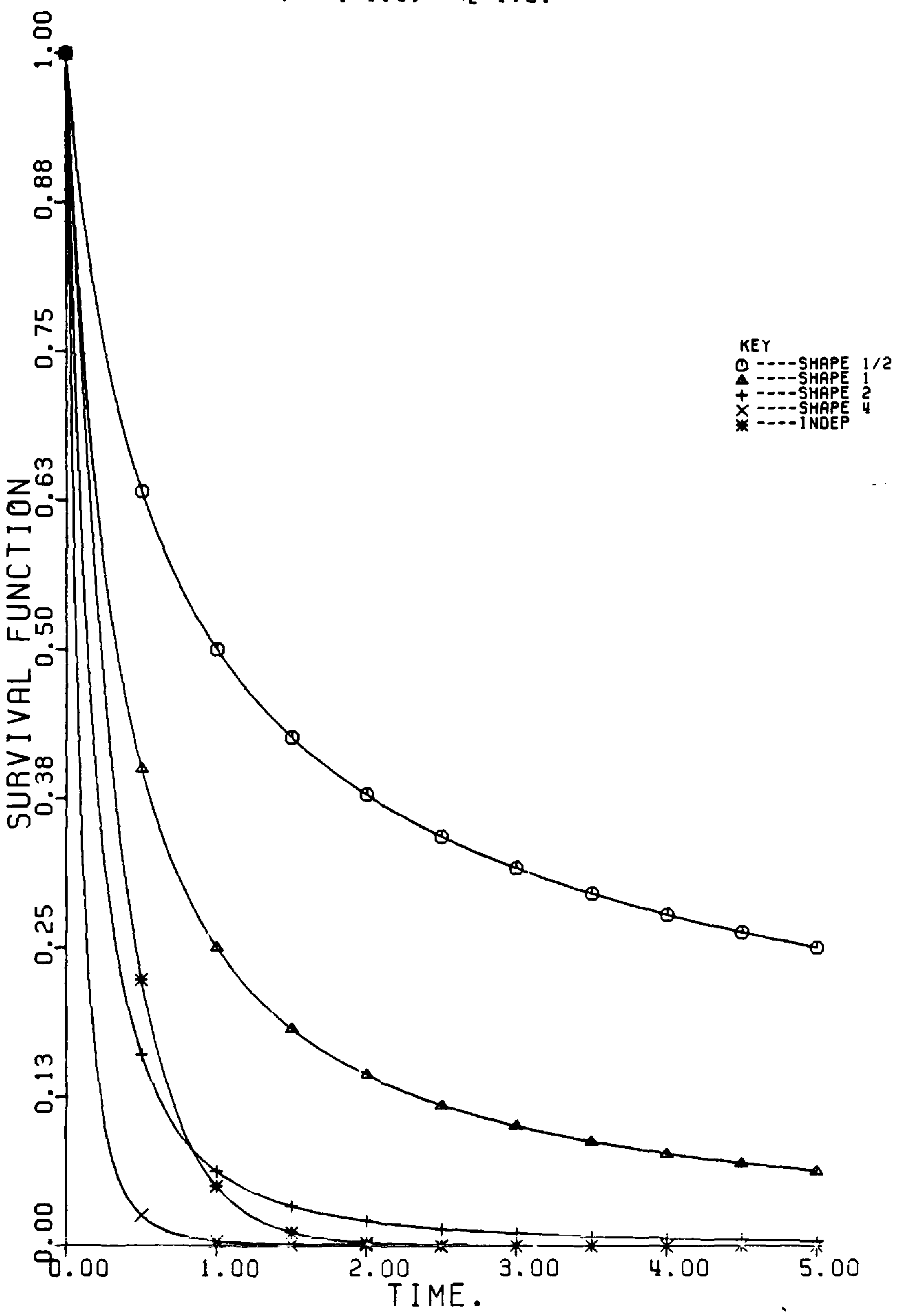
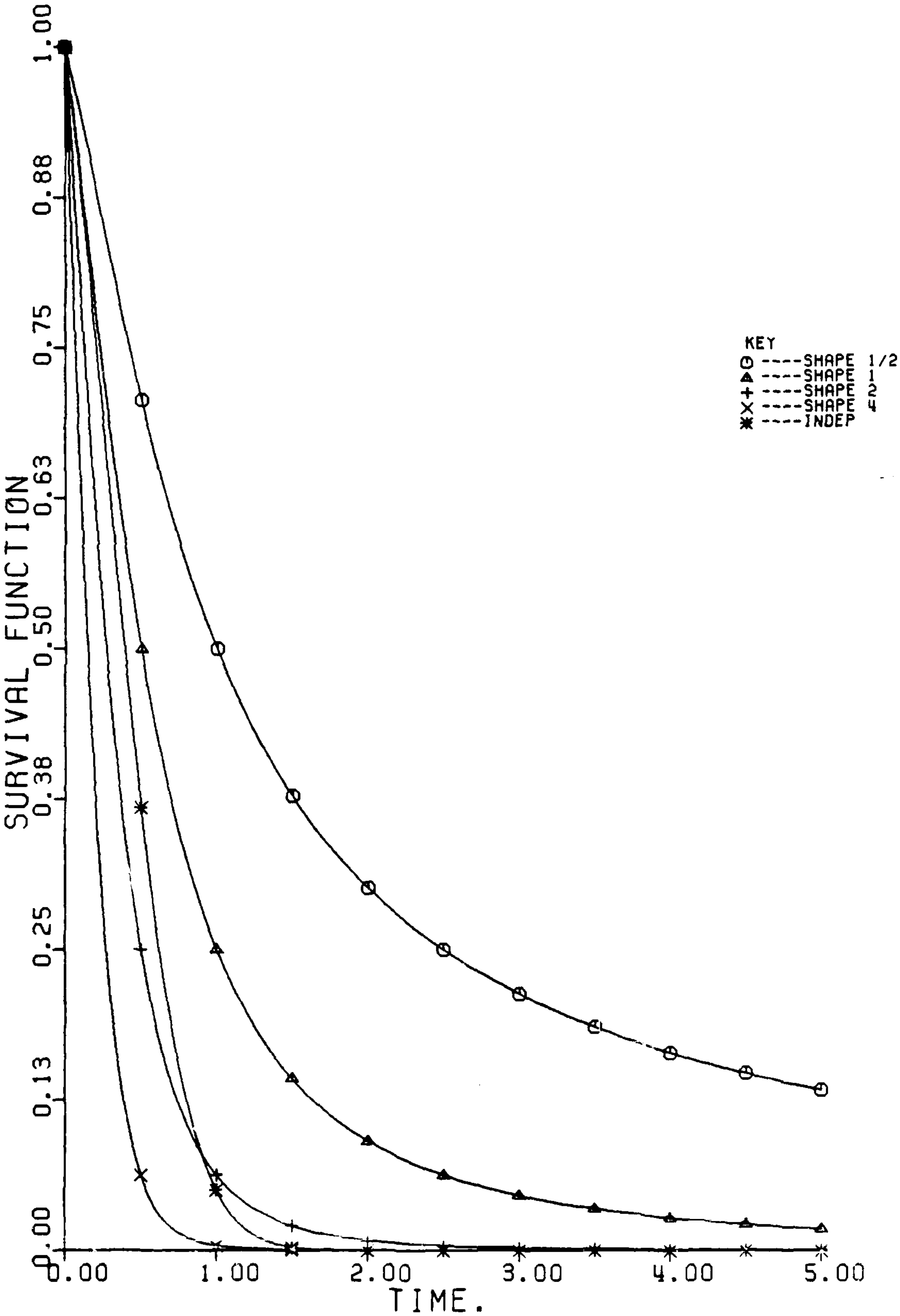


FIGURE 2 B

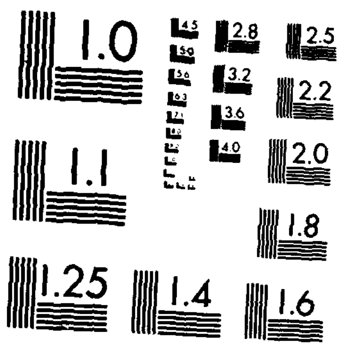
SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0,\quad \lambda_2=2.0,\quad \eta_1=1.0,\quad \eta_2=2.0.$



AD-A174 825 EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES IF THEY ARE ACTUA (U) OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS M L MOESCHBERGER ET AL 26 NOV 85 2/2

UNCLASSIFIED AFOSR-TR-86-2042 AFOSR-82-0307 F/G 14/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

FIGURE 2 C

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=2.0$, $\eta_2=2.0$.

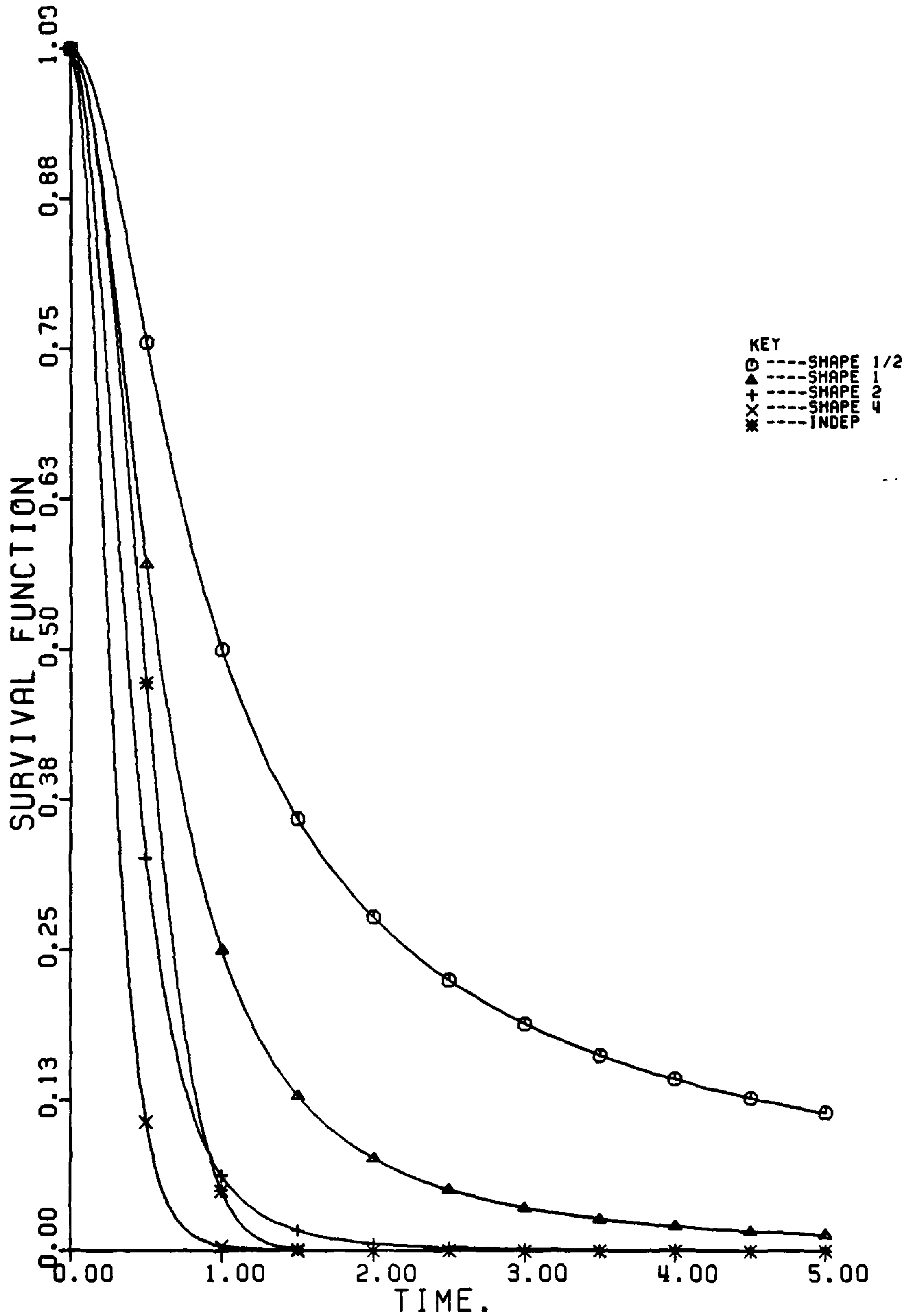


FIGURE 2 D

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1/2$, $\eta_2=1/2$.

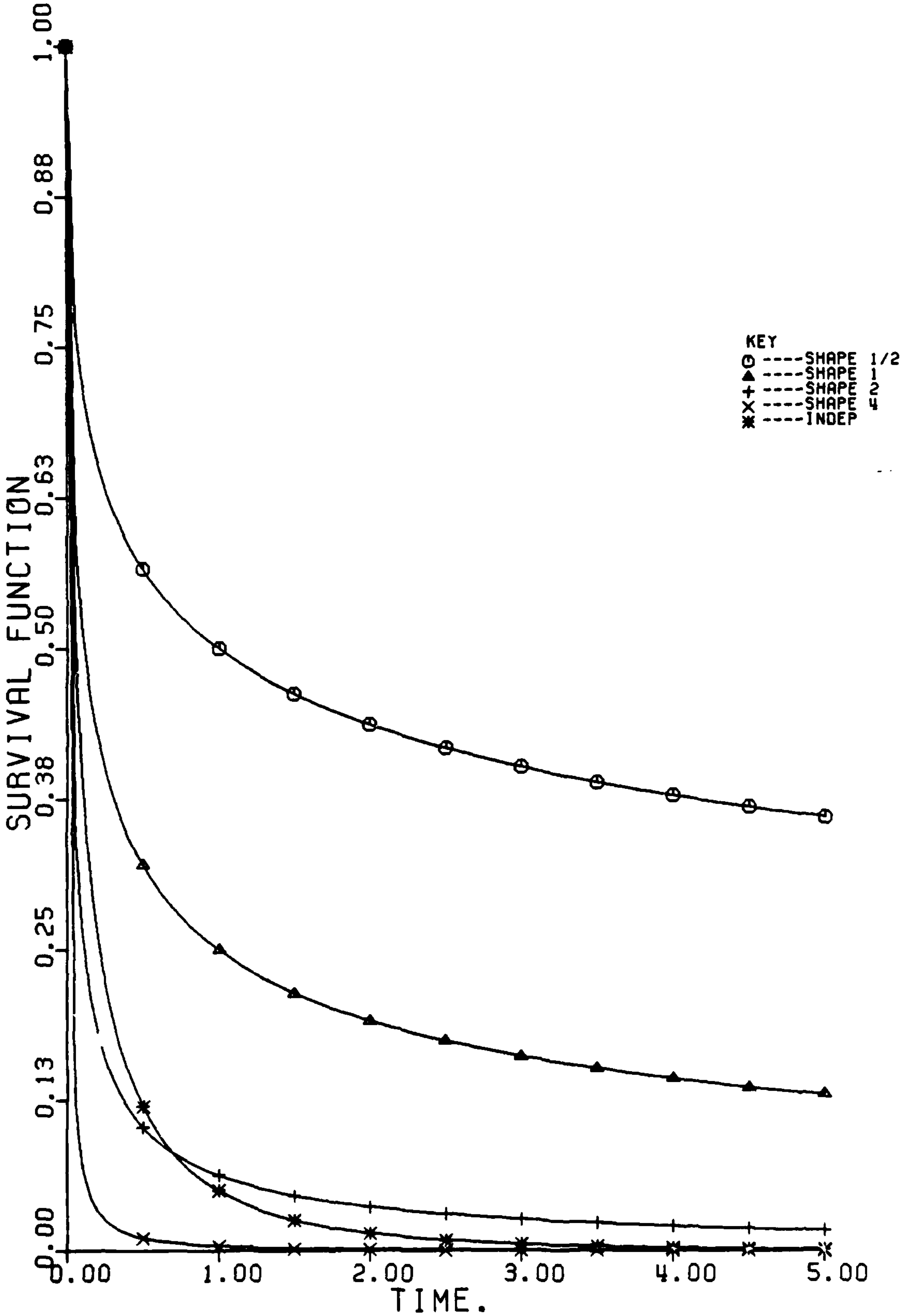


FIGURE 2 E

SERIES SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=2.0.$

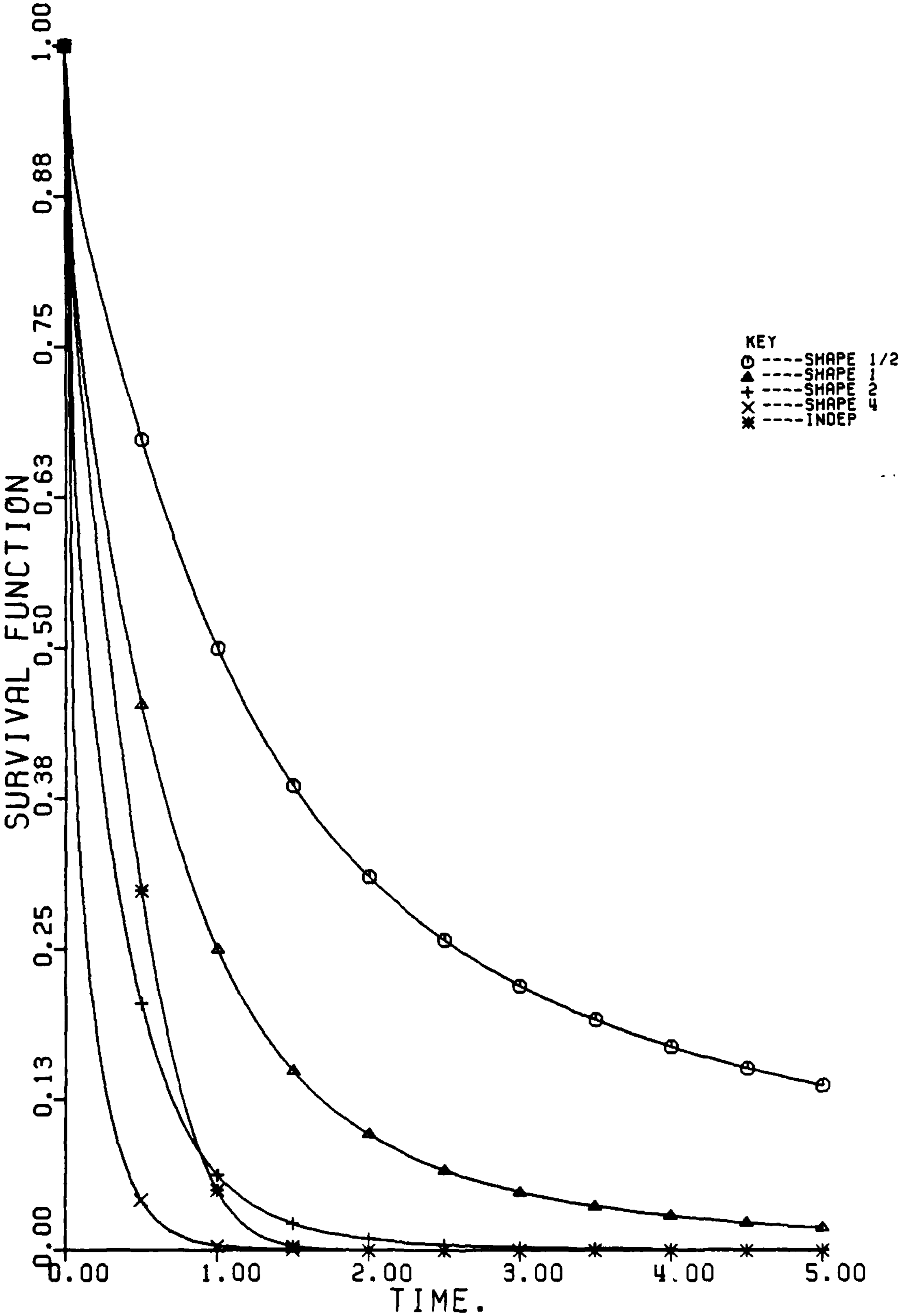


FIGURE 3 A

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=1.0.$

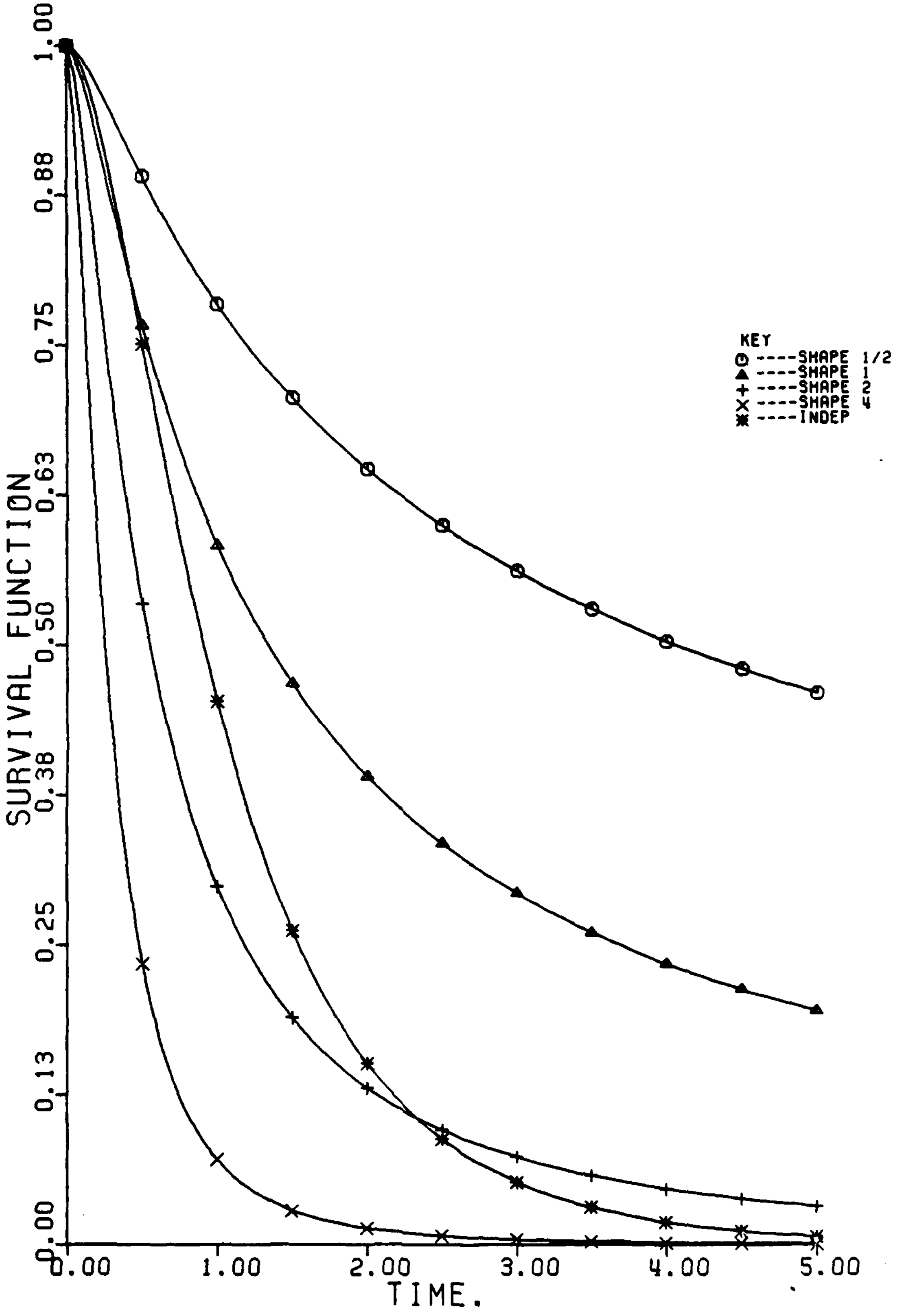


FIGURE 3 B

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=2.0.$

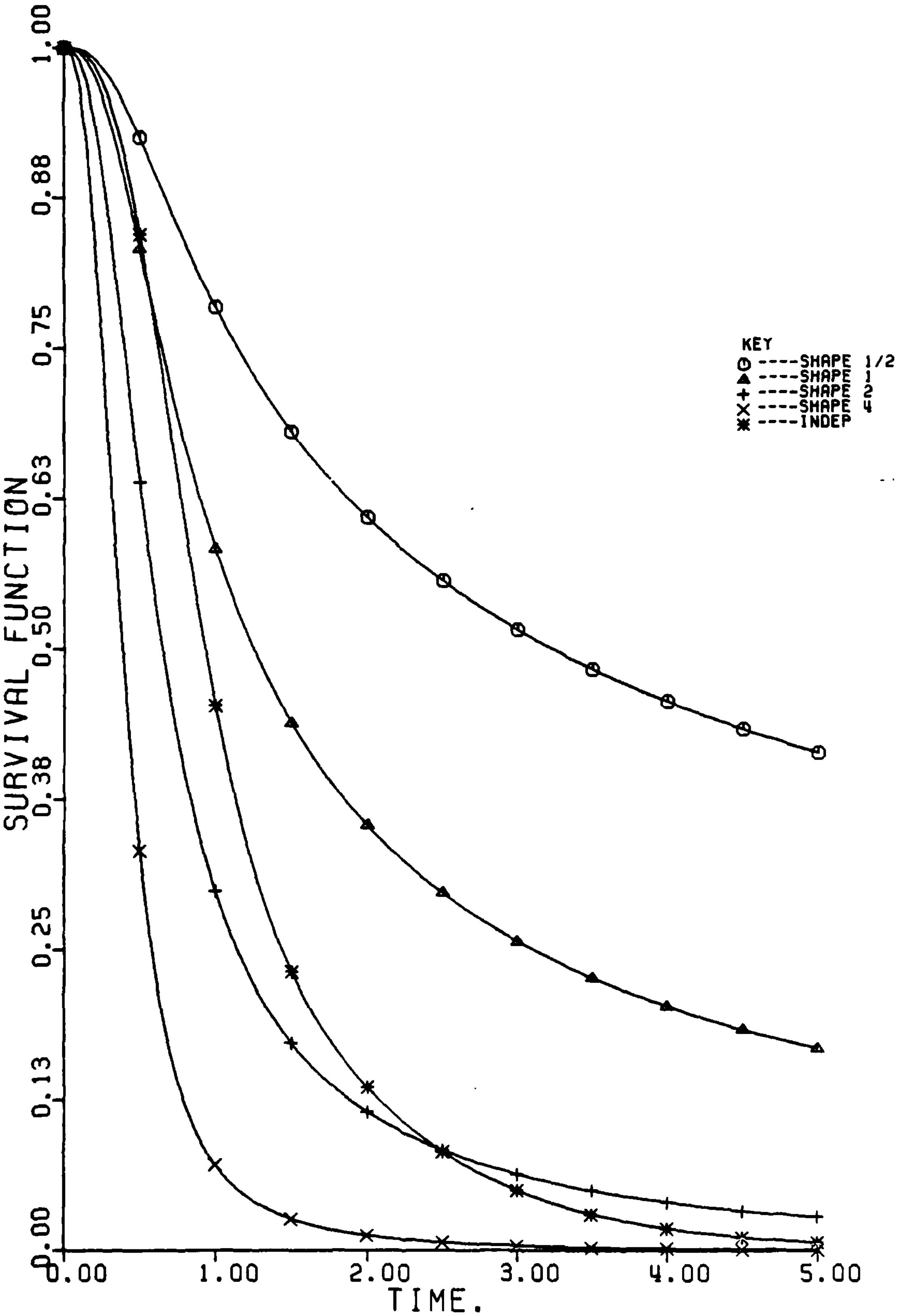


FIGURE 3 C

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=2.0$, $\eta_2=2.0$.

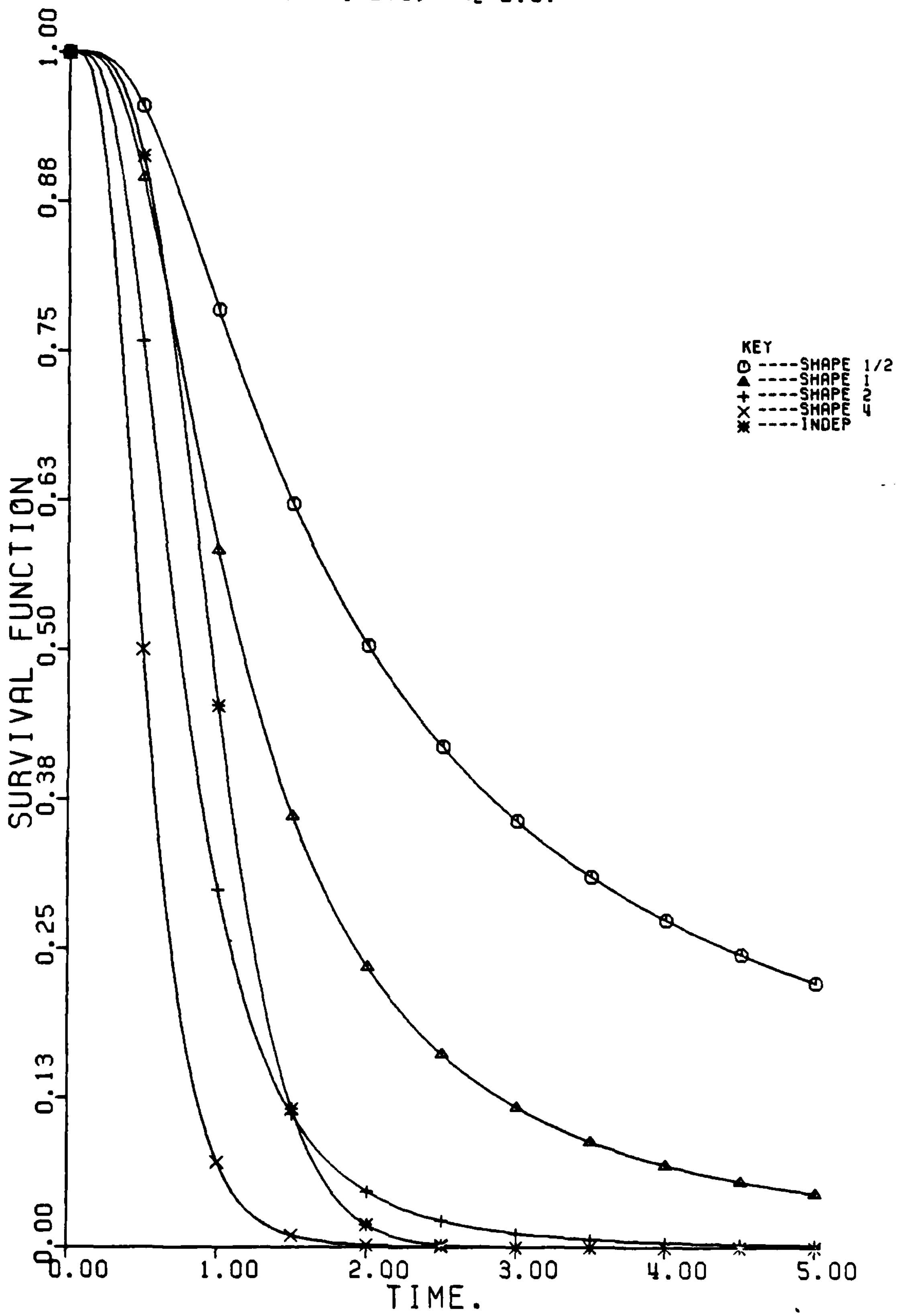
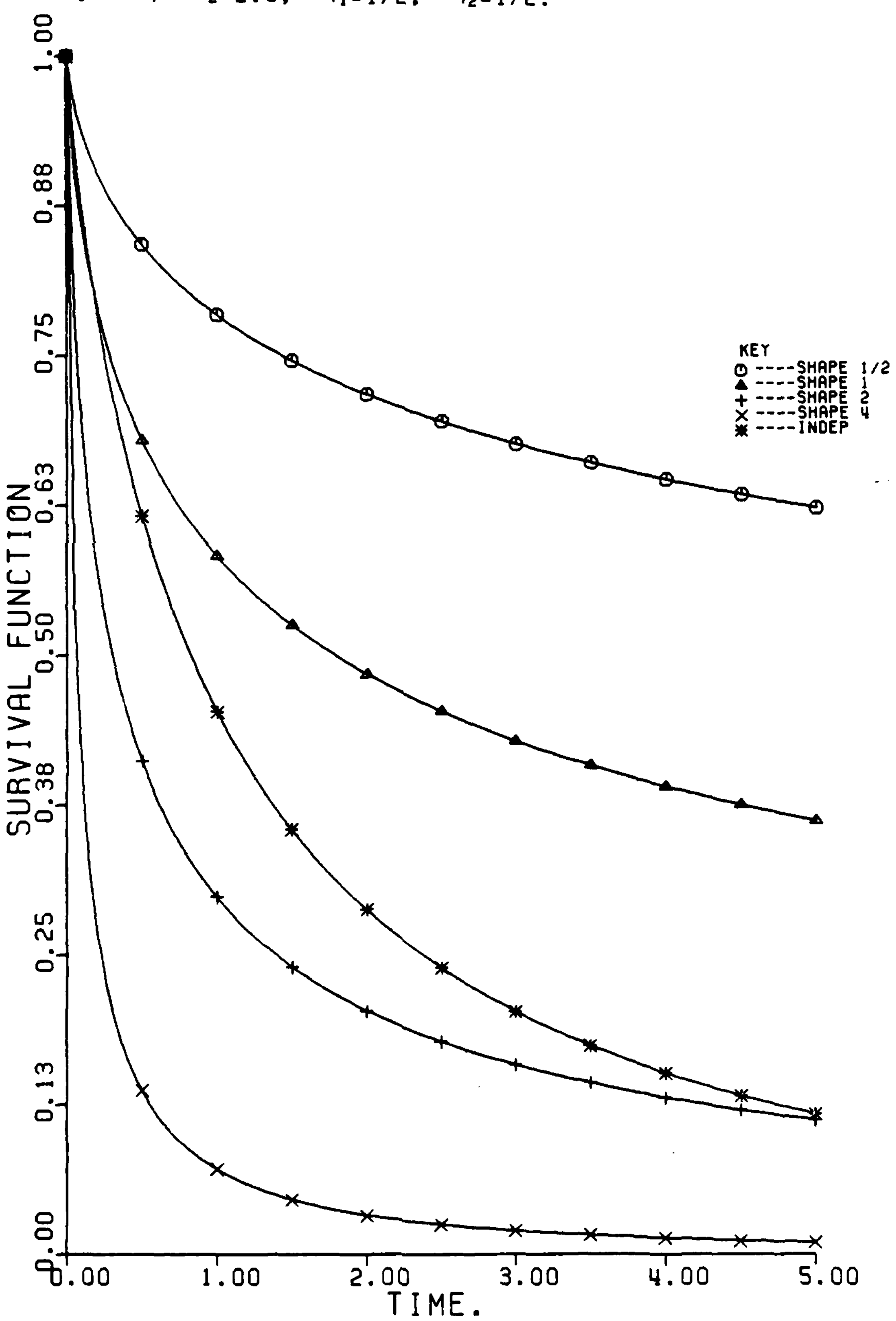


FIGURE 3 D

PARALLEL SYSTEM RELIABILITY UNDER GAMMA(A,1) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1/2$, $\eta_2=1/2$.



and

$$\mathrm{Cov}(X,Y) = \frac{\Gamma(1+1/\eta_1)\,\Gamma(1+1/\eta_2)}{\lambda_1^{1/\eta_1}\lambda_2^{1/\eta_1}}\Bigg\{\frac{\eta_1\eta_2}{(\eta_1\eta_2-\eta_1-\eta_2)}\frac{\left(b^{\frac{\eta_1\eta_2-\eta_1-\eta_2}{\eta_1\eta_2}} - a^{\frac{\eta_1\eta_2-\eta_1-\eta_2}{\eta_1\eta_2}}\right)}{(b-a)} -$$

$$\frac{\eta_1\eta_2}{(\eta_1-1)(\eta_2-1)}\frac{\left(b^{\frac{\eta_1-1}{\eta_1}} - a^{\frac{\eta_1-1}{\eta_1}}\right)\left(b^{\frac{\eta_2-1}{\eta_2}} - a^{\frac{\eta_2-1}{\eta_2}}\right)}{(b-a)^2}\Bigg\} \text{if } \eta_1 \neq 1,\ \eta_2 \neq 1,\ 1/\eta_1 + 1/\eta_2 \neq 1$$

$$\frac{\Gamma(1+1/\eta_1)\,\Gamma((2\eta_1-1)/\eta_1)}{\lambda_1^{1/\eta_1}\lambda_2^{\eta_1/(\eta_1-1)}}\left[\frac{\ln(b/a)}{(b-a)} - \frac{\eta_1^2}{(\eta_1-1)}\frac{\left(b^{\frac{\eta_1-1}{\eta_1}} - a^{\frac{\eta_1-1}{\eta_1}}\right)}{(b-a)^2}\left(b^{1/\eta_1} - a^{1/\eta_1}\right)\right]$$

$$\text{if } 1/\eta_1 + 1/\eta_1 = 1$$

$$\frac{\Gamma(1+1/\eta_i)}{\lambda_i^{1/\eta_i}\lambda_{i'}}\left[\frac{\eta_i\eta_{i'}}{(\eta_i\eta_{i'}-\eta_i-\eta_{i'})(b-a)}\left(b^{\frac{\eta_i\eta_{i'}-\eta_i-\eta_{i'}}{\eta_i\eta_{i'}}} - a^{\frac{\eta_i\eta_{i'}-\eta_i-\eta_{i'}}{\eta_i\eta_{i'}}}\right) - \frac{\eta_i}{(\eta_i-1)}\frac{\left(b^{\frac{\eta_i-1}{\eta_i}} - a^{\frac{\eta_i-1}{\eta_i}}\right)\ln(b/a)}{(b-a)^2}\right]$$

$$\text{if } \eta_i \neq 1,\ \eta_{i'} = 1$$

$$\frac{1}{(\lambda_1\ \lambda_2)}\left[\frac{1}{(ab)} - \frac{\ln(b/a)^2}{(b-a)^2}\right] \text{ if } \eta_1 = \eta_2 = 1$$

For this model, the reliability function for a series system is

$$R_s(t) = \underline{[\exp(-b(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2}) - \exp[-a(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})]]} \qquad (2.7)$$

$$(b-a)(\lambda_1 t^{\eta_1} + \lambda_2 t^{\eta_2})$$

and for a parallel system is

$$R_p(t) = \frac{[\exp(-b(\lambda_1 t^{\eta_1}) - \exp(-a\lambda_1 t^{\eta_1})}{(b-a)\lambda_1 t^{\eta_1}} + \frac{[\exp(-b\lambda_2 t^{\eta_2}) - \exp(-a\lambda_2 t^{\eta_2})}{(b-a)\lambda_2 t^{\eta_2}} - Rs(t)$$

Figures 4A-E show the reliability for a series system and figures 5A-E for a parallel system under the uniform model for various combinations of $\lambda_1, \lambda_2, \eta_1, \eta_2$ a,b. Notice that when A = .25, B = .75, which corresponds to an operating environment which is less severe than the test environment, the system reliability is greater than that expected under independence, while when (a,b) = (1.25, 1.75) or (1., 2), which corresponds to an environment more severe than the test environment, the system reliability is smaller. Also when the (a,b) contains 1, which corresponds to an environment which incurs the possibility of no differential effect from that found in the laboratory, there is little difference in the dependent and independent system reliability.

3. Estimation of Parameters Under Gamma Model

Consider the model (2.3) with $\eta_1 = \eta_2 = \eta$. For this model, the reliablity for a series system is

$$R_s(t) = \left(1 + \frac{(\lambda_1+\lambda_2)}{b} t^{\eta}\right)^{-a}. \tag{3.1}$$

Notice that this model depends only on two parameters $\theta = (\lambda_1 + \lambda_2)/b$ and a so that if we had data only from systems on test in the operating environment, the only identifiable parameters are a, θ,

FIGURE 4 A

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=1.0.$

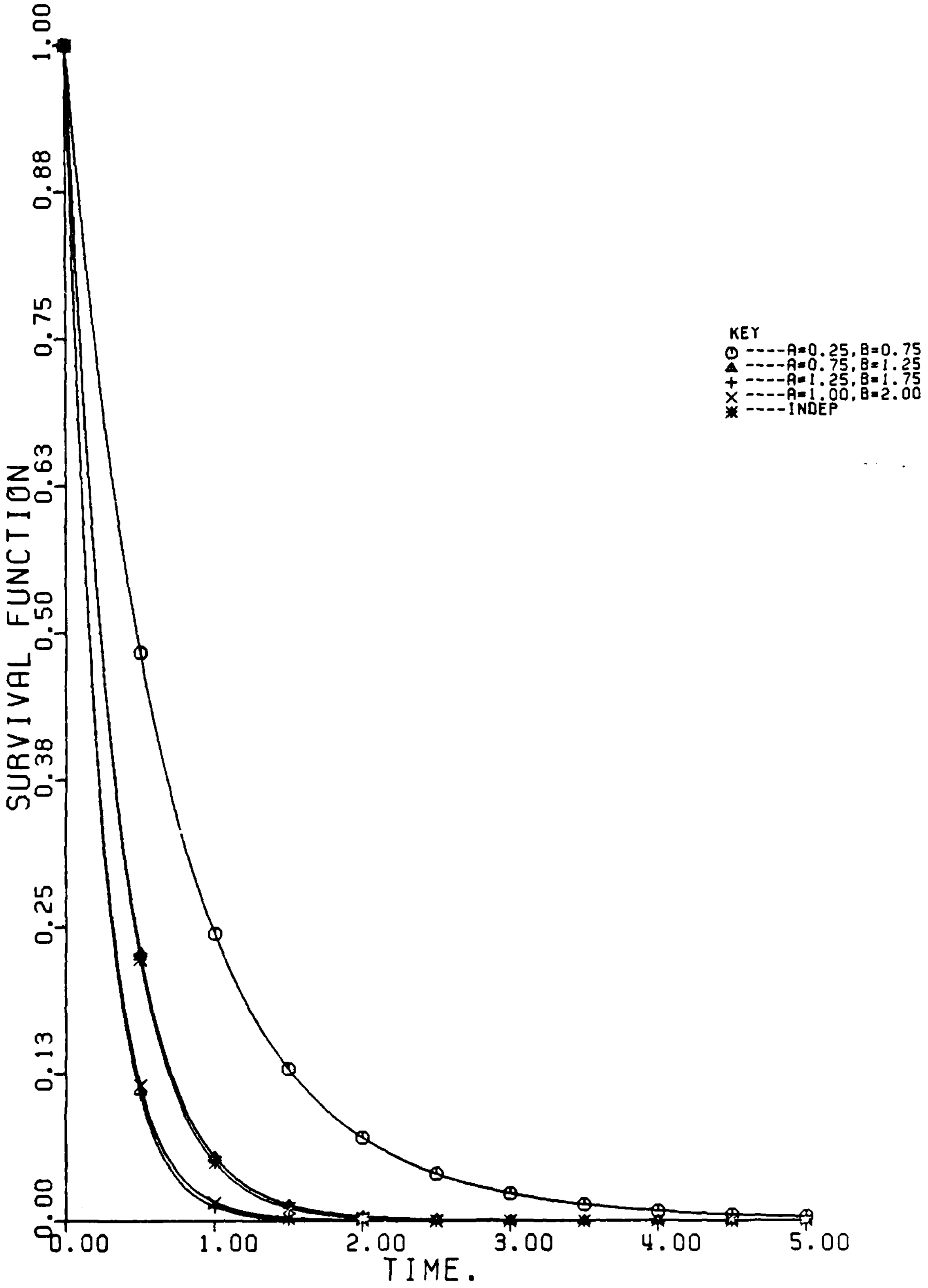


FIGURE 4 B

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1.0, \quad \eta_2=2.0.$

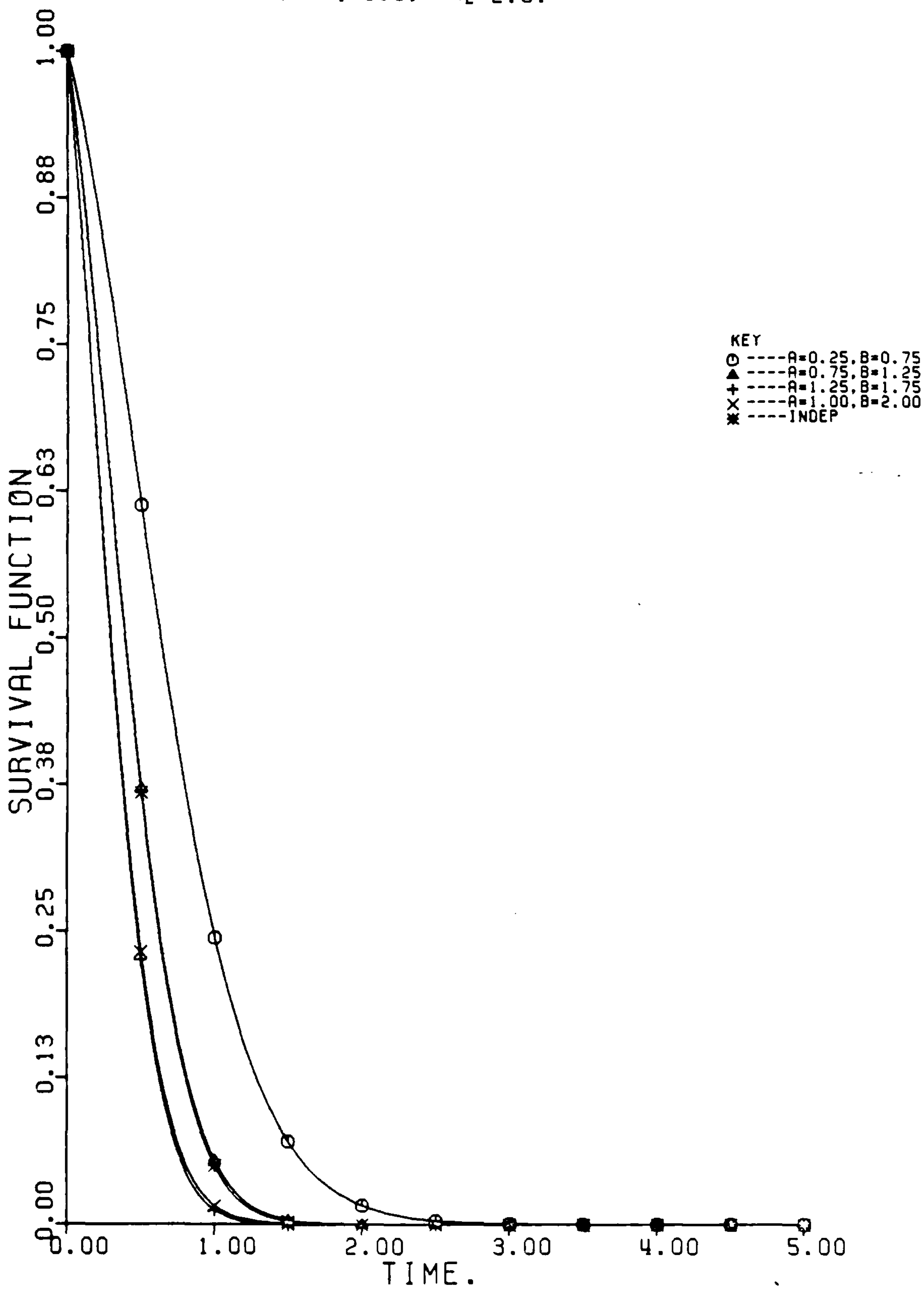


FIGURE 4 C

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=2.0, \quad \eta_2=2.0.$

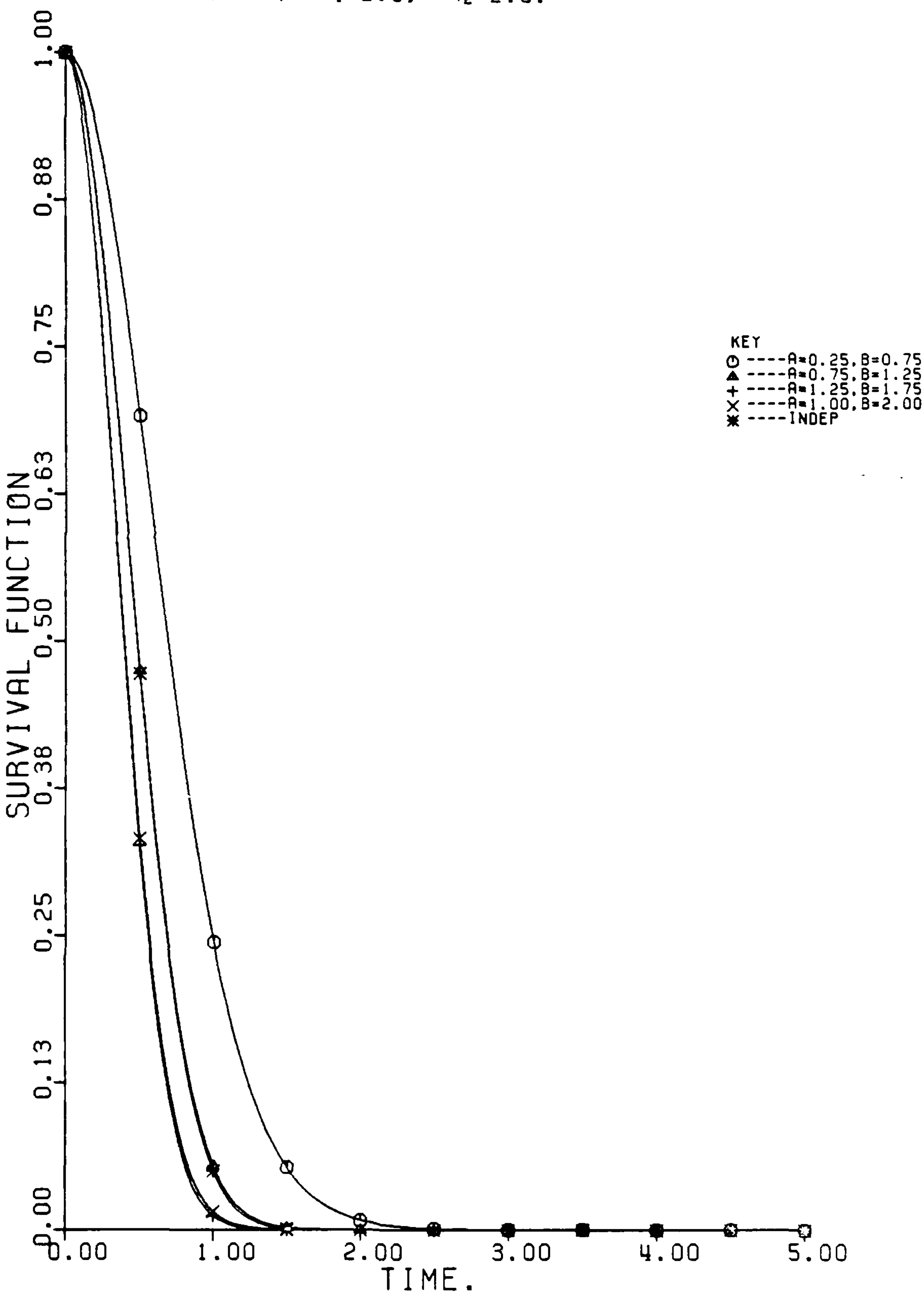


FIGURE 4 D

SERIES SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2, \quad \eta_2=1/2.$

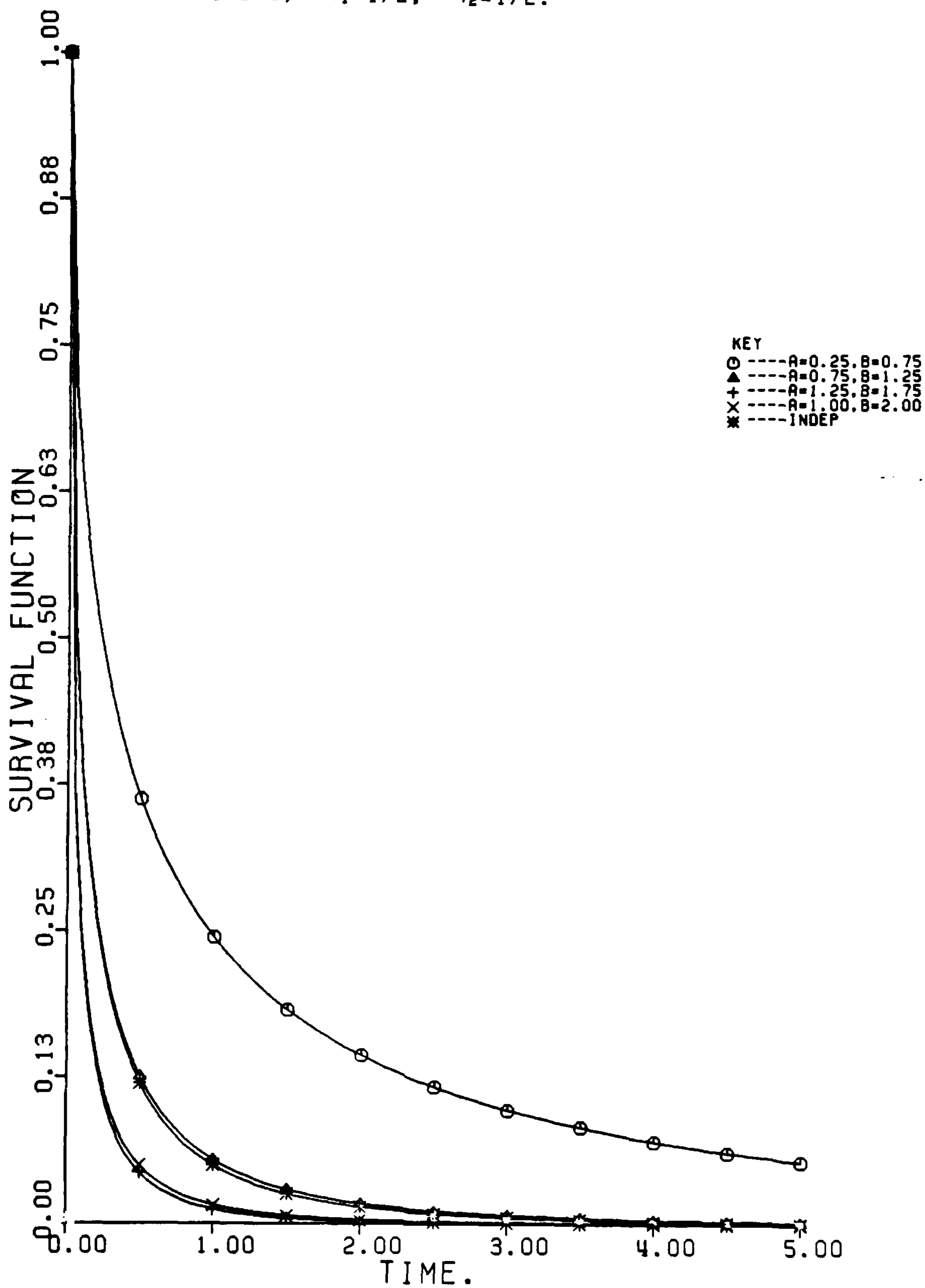


FIGURE 5 A

PARALLEL SYSTEM RELIABILITY UNDER UNIF(A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=1.0$.

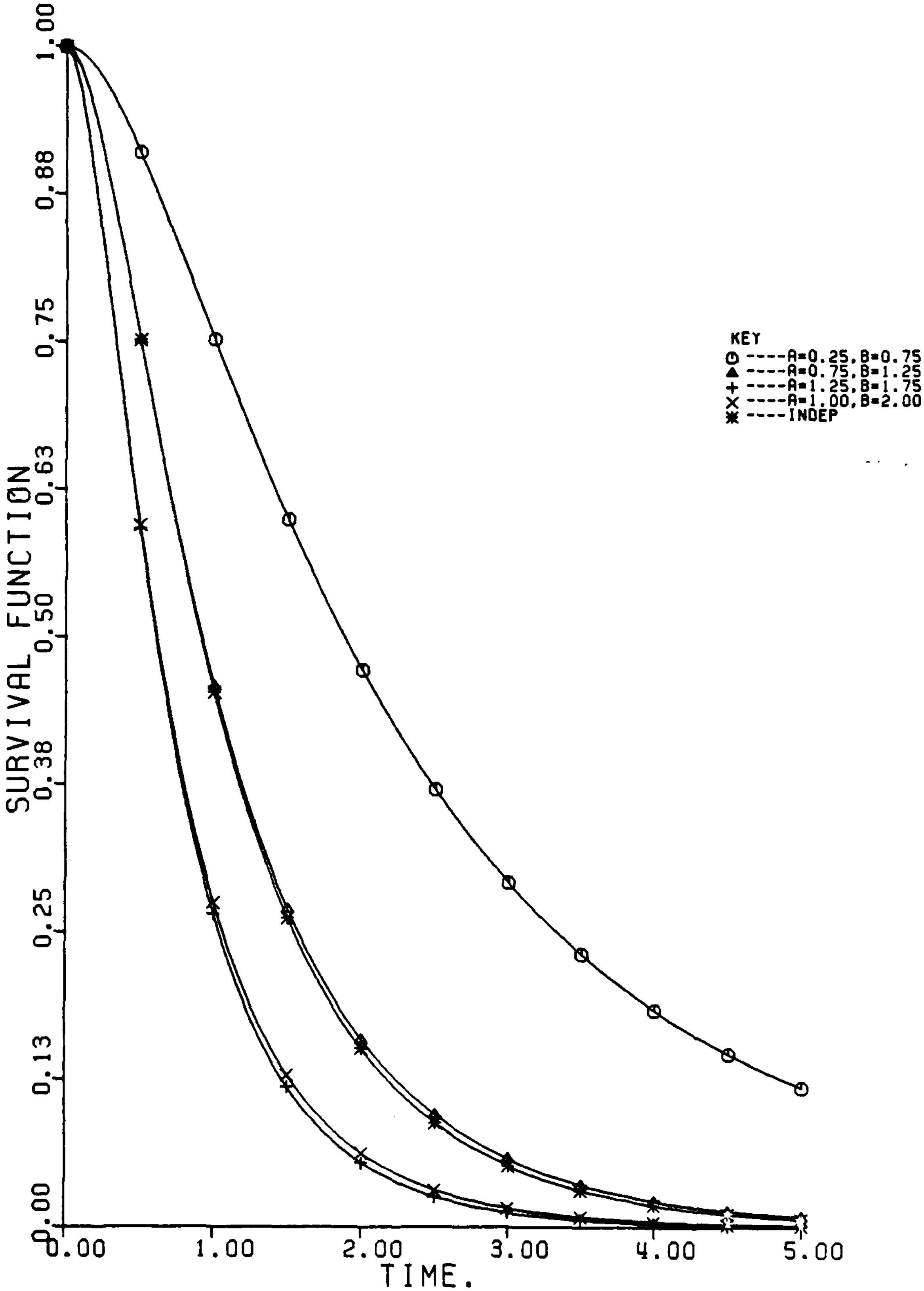


FIGURE 5 B

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1.0$, $\eta_2=2.0$.

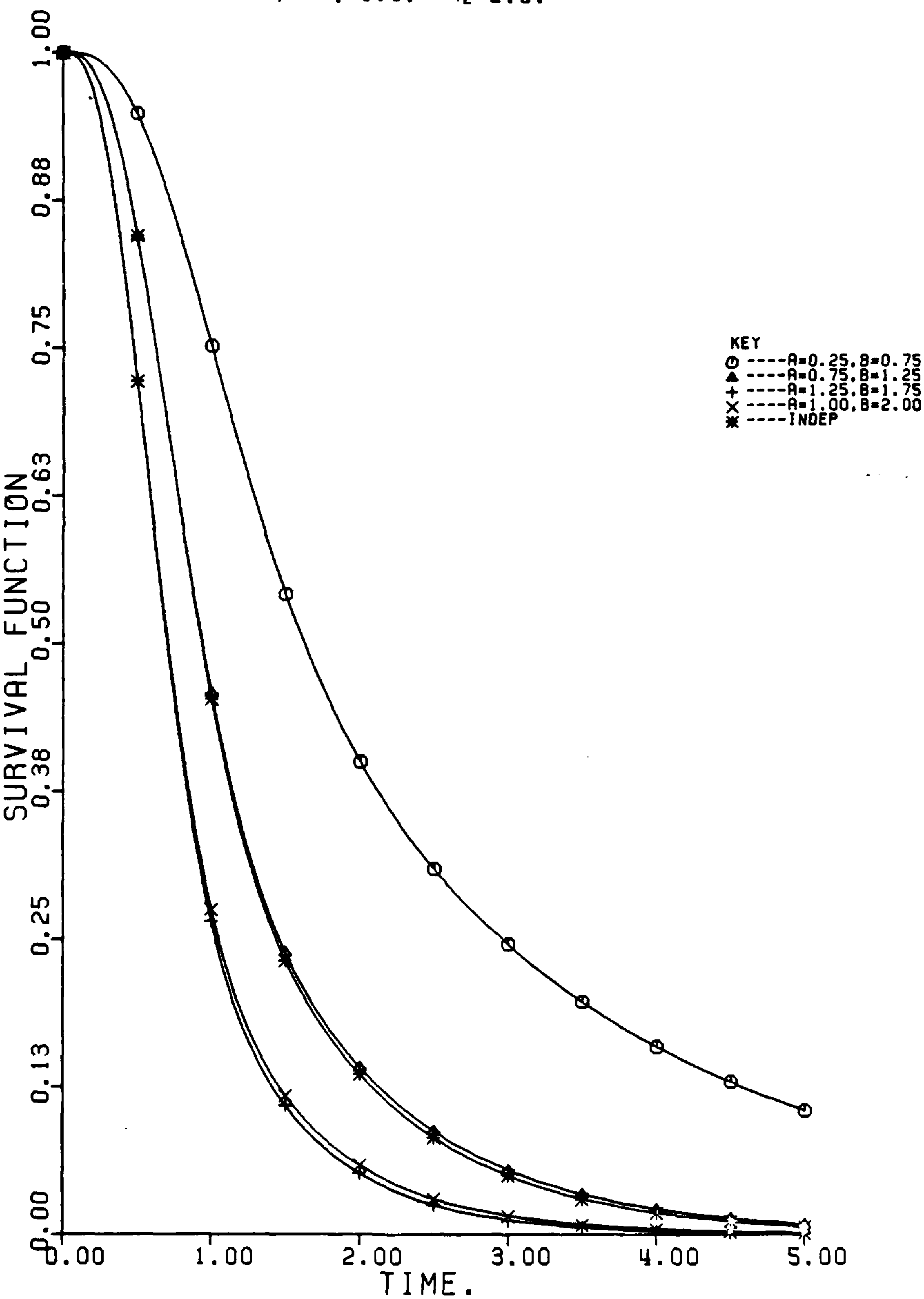


FIGURE 5 C

PARALLEL SYSTEM RELIABILITY UNDER UNIF(A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=2.0$, $\eta_2=2.0$.

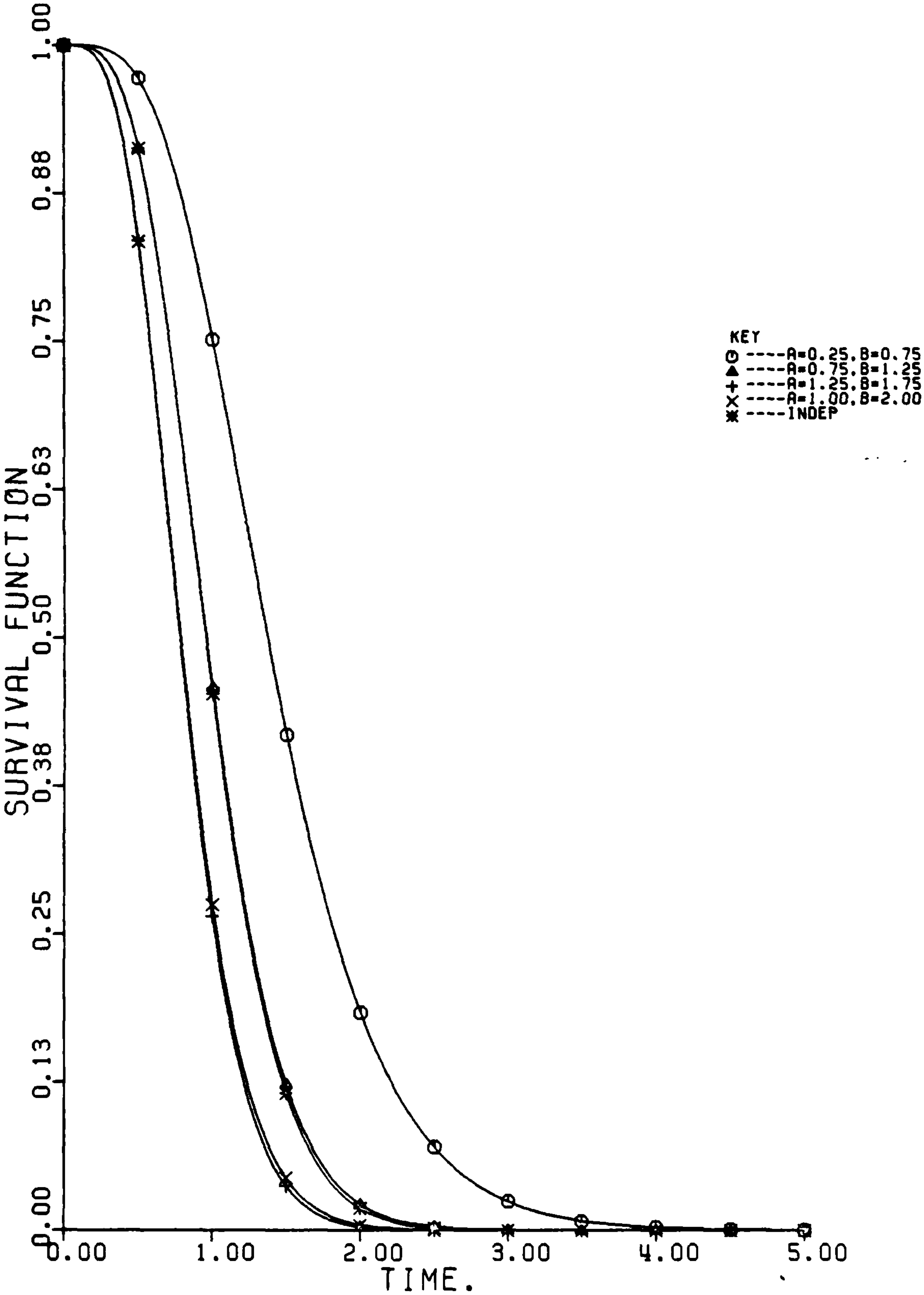


FIGURE 5 D

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0$, $\lambda_2=2.0$, $\eta_1=1/2$, $\eta_2=1/2$.

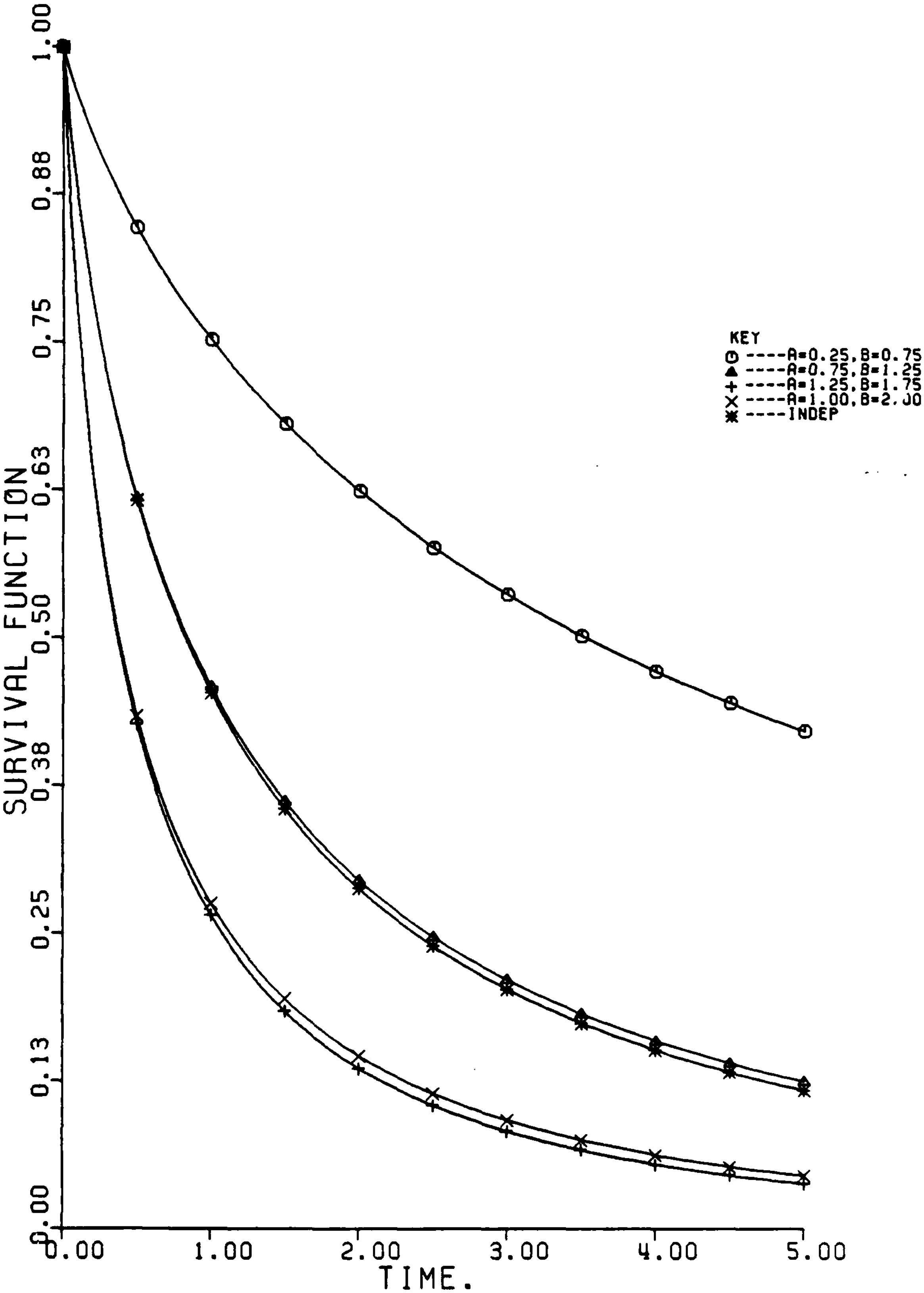
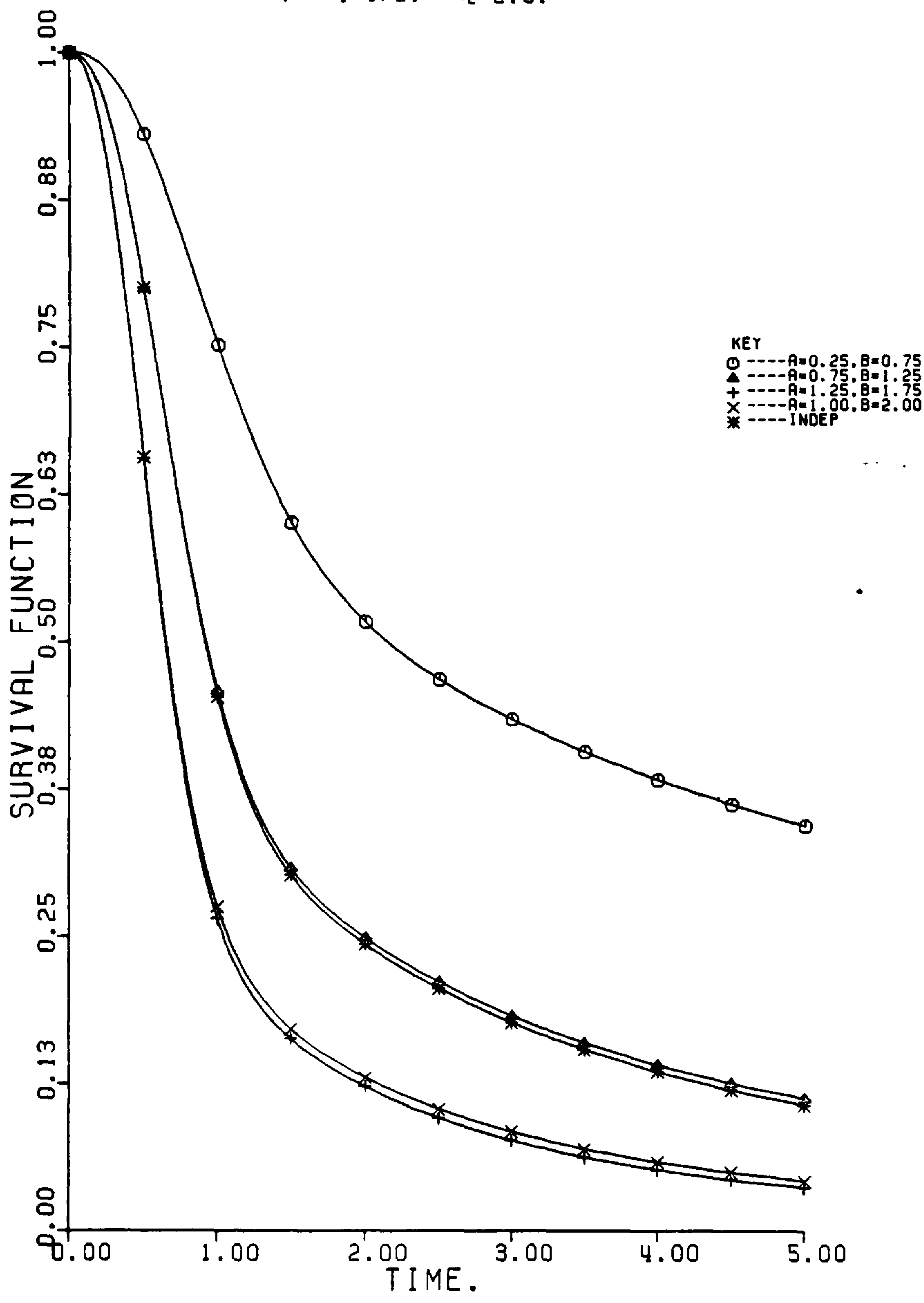


FIGURE 5 E

PARALLEL SYSTEM RELIABILITY UNDER UNIF (A,B) MODEL FOR THE ENVIRONMENTAL STRESS.

$\lambda_1=1.0, \quad \lambda_2=2.0, \quad \eta_1=1/2. \quad \eta_2=2.0.$



η, not $\lambda_1, \lambda_2, \eta$, a, b. However, in many instances we have extensive data on the performance of the components in the lab under ideal operating conditions so that one may consider $\lambda_1, \lambda_2, \eta$ to be known based on estimates from this data. We shall focus on the problem of estimating θ and a, based on data on the system failure times collected in the operating environment. Let $t_1, \ldots, t_n$ be the failure times for n such systems put on test, and, let $w_i = t_i^{\eta}$, $i = 1, \ldots, n$.

Prior to attempting to estimate (a, θ), we would like to check if the model (3.1) is feasible. A graphical check of this model can be done through the scaled total time on test (STTOT) plot of Barlow and Campo (1975). The STTOT for W is

$$G_W(t) = \frac{\int_0^{F^{-1}(t)} Rs(t)dt}{\int_0^{F^{-1}(1)} Rs(t)dt} = 1-(1-t)^{(a-1)/a} \text{ for } a > 1. \tag{3.2}$$

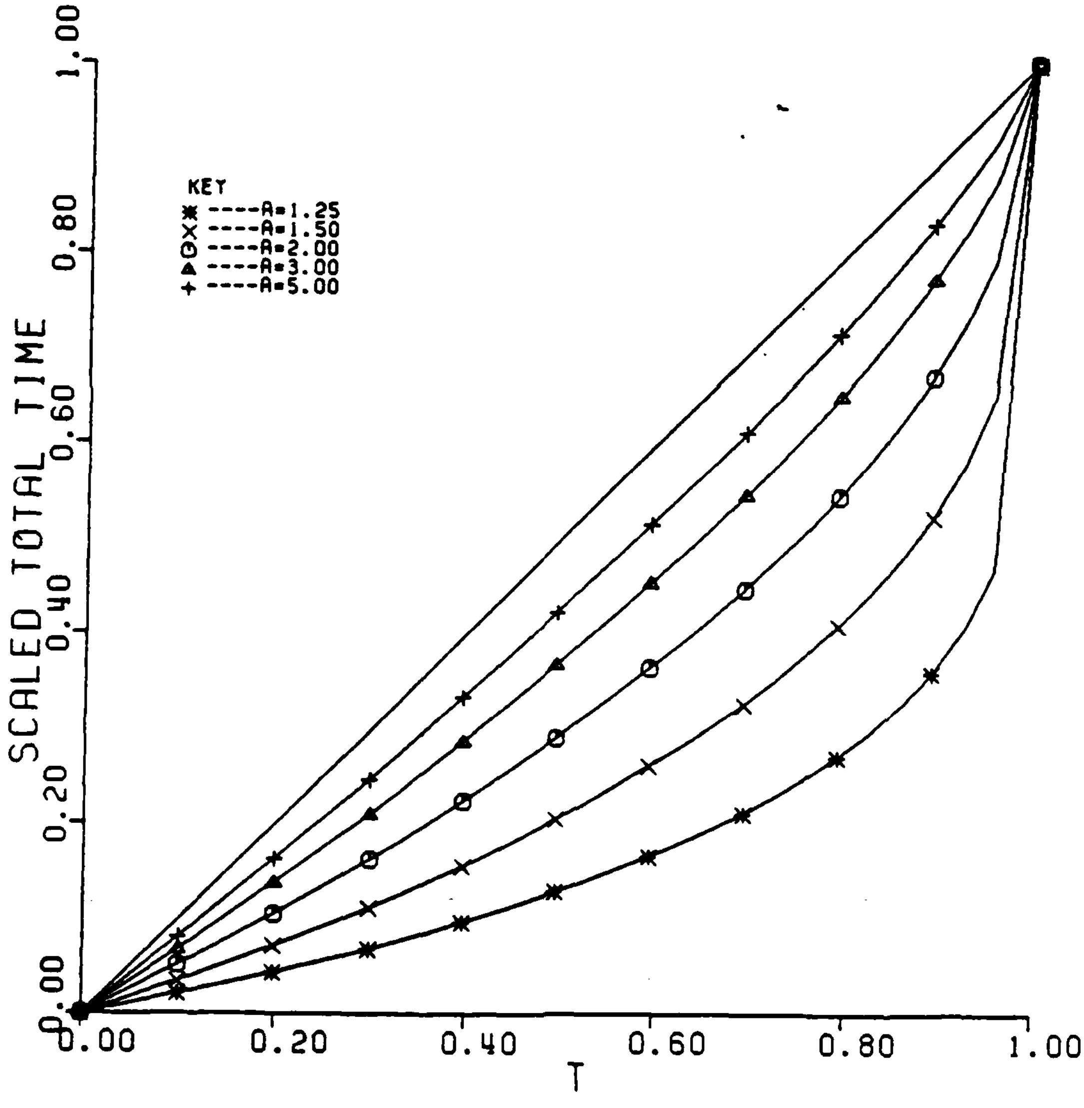
Note that (3.2) depends only on a. Figure 6 shows the form of the STTOT for several values of a. Notice that for all a, the STTOT is below the 45° line (which corresponds to exponential system life) since the hazard rate of the series system is decreasing. Let

$$Tn(W_{(i)}) = \sum_{j=1}^{i} W_{(j)} + (n-i)W_{(i)}, \tag{3.3}$$

where $W_{(1)} \leq W_{(2)} \leq \ldots \leq W_{(n)}$ are the ordered systems failure times be the total time on test at $W_{(i)}$. The empirical STTOT plot then plots $(i/n, T_n(W_{(i)})/T(W_{(n)}))$ which can be compared to figure 6 for a graphical check of the model. Also, crude estimates of a can be obtained by

FIGURE 6

SCALED TOTAL TIME ON TEST TRANSFORM FOR GAMMA MODEL.



comparing the empirical and theoretical STTOT plots. When there is no random environmental effect and the components are independent, then the empirical STTOT plot should look like the 45° line. Also as a tends to infinity this plot approaches the 45° line.

We now consider several estimates of a and θ. The log likelihood for the model (3.1), based on a sample of size n, is

$$L(a,\theta) = n \ln a + n \ln \theta - (a+1)\sum_{i=1}^{n} \ln (1+\theta W_i) \tag{3.4}$$

so that

$$\partial/\partial a \, L(a,\theta) = n/a - \sum_{i=1}^{n} \ln (1 + \theta W_i) \tag{3.5}$$

$$\text{and } \partial/\partial\theta \, L(a,\theta) = n/\theta - (a+1) \sum_{i=1}^{n} w_i/(1+ \theta w_i) \tag{3.6}$$

For (3.5) we note that the maximum likelihood estimator of a

$$a_{mle} = \frac{n}{\sum_{i=1}^{n} \ln (1 + \theta W_i)} \tag{3.7}$$

and the maximum likelihood estimator of θ is the solution to

$$\frac{n}{\theta} - \left(\frac{n}{\sum \ln (1 + \theta W_i)} + 1\right) \left(\sum \frac{W_i}{1 + \theta W_i}\right) = 0. \tag{3.8}$$

One can show that θ is positive if $n \sum_{i=1}^{n} w_i^2 > 2(\sum_{i=1}^{n} w_i)^2$. (3.9)

In such case θ_{mle} is obtained by solving

(3.8) numerically.

A second estimator of (a, θ) is the method of moments (mme). Since $E(W) = [\theta(a-1)]^{-1}$ and $E(W^2) = 2[\theta^2(a-1)(a-2)]^{-1}$ where $a > 2$, we have

$$a_{mme} = 1 + \frac{\Sigma w_i^2}{\Sigma w_i^2 - 2(\Sigma w_i^2)} \quad (3.10) \text{ and } \theta_{mme} = \frac{\Sigma w_i^2 - 2(\Sigma w_i)^2}{\Sigma w_i (\Sigma w_i)^2} \qquad (3.11)$$

provided that (3.9) holds. If (3.9) does not hold, then this estimator does not exist.

A third estimator was suggested by Berger (1983) in a different context. He suggested estimating θ a modified methods of moments estimator $\theta_{ber} = (a\,w)^{-1}$, (3.12)

where $w = \Sigma w_i/n$,

which is used as the true value of θ in the likelihood (3.4) so that the estimator of a is the solution to

$$-\Sigma\, \ell n\left(1+\frac{w_i}{aw}\right) + \frac{(a+1)}{wa^2}\Sigma \frac{w_i}{1+w_i/(aw)} = 0 \qquad (3.13)$$

A final estimator is based on the STTOT plot. Let $C_i = \ell n(1-i/n)$ and $D_i = \ell n(1-Tn(W_{(i)})/Tn(W_{(n)}))$, $i = 1, \ldots, n-1$. If (3.2) holds, then we should have

$$D_i \approx (1-1/a)C_i, \; i = 1, \ldots, n-1, \qquad (3.14)$$

so the value of a which minimizes

$\sum_{i=1}^{n-1} (D_i - (1-1/a)\, C_i)^2$ is a reasonable estimator of a

The resulting estimator is $a_{ls} = \dfrac{\Sigma C_i^2}{\Sigma C_i^2 - \Sigma C_i D_i}$ (3.15) which is in the parameter space if $\Sigma C_i^2 > \Sigma C_i D_i$. A better estimator should be obtained by weighting the D_i's differently since for $i < j$, $\text{Var}(D_i) < \text{Var}(D_j)$. The variance of D_i depends on the unknown parameter a so we weight by the variance of D_i computed under an assumed exponential distribution. The variance of D_i in that case is

$$V_i = \sum_{j=1}^{i} \frac{1}{(n-j)^2}, \; i = 1, \ldots, n-1 \tag{3.16}$$

so that the weighted least squares estimator of a is

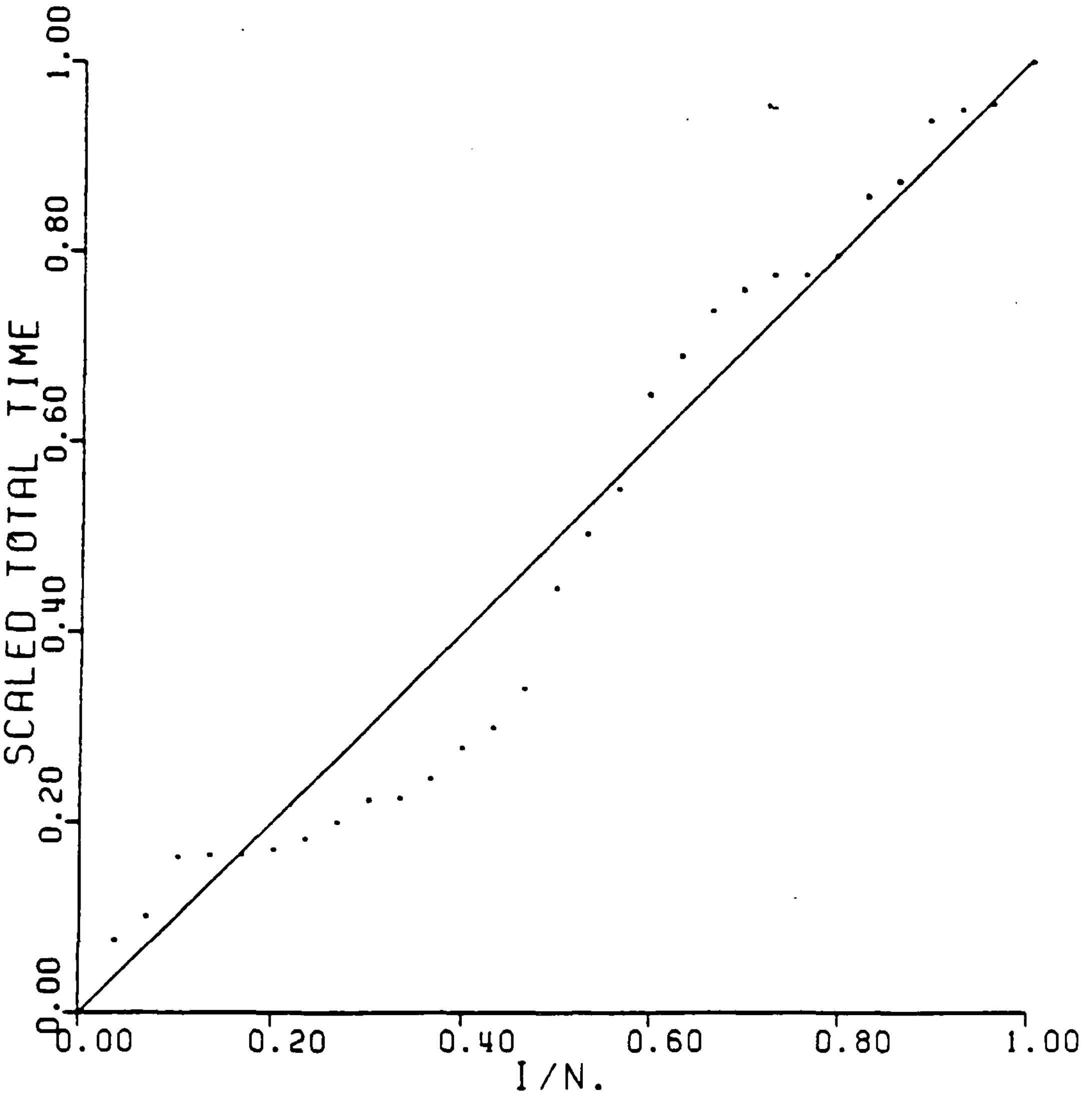
$$a_{wls} = \frac{\Sigma C_i^2/V_i}{\left(\Sigma \dfrac{C_i^2}{V_i} - \Sigma \dfrac{C_i D_i}{V_i}\right)} \text{ if } \Sigma C_i^2/V_i > \Sigma C_i D_i/V_i. \tag{3.17}$$

Once we have obtained a by either of the two least squares estimators, we substitute this value into (3.6) and solve this equation numerically for θ_{ls} or θ_{wls}.

The condition $\Sigma C_i^2/V_i > \Sigma C_i D_i/V_i$ includes a few more possible samples than the condition (3.9) for the other three estimators. However, those samples which satisfy $\Sigma C_i^2/V_i > \Sigma C_i D_i/V_i$ for which (3.9) fails to obtain yield very large estimates of θ. Since a reasonable model for T when θ and a are not estimable is the independent Weibull series system which has system reliability very close to (3.1) when a is very large, this is not a problem. Figures 7a and 7b are scaled total time on test plots from two simuilated samples of size 30 from (3.1) with $a = 3$, $\theta = 1$. Looking at figure 7a, we see that the estimated scaled total on test doesn't look too different from the 45° so that an exponential model might not be unreasonable. For this data set only the weighted least squares estimator exists and it yields $a_{WLS} = 45.33$ and $\theta = .0567$. For the data in figure 7b all estimates exist, and we have

FIGURE 7 A

SCALED TOTAL TIME ON TEST PLOT
FOR SIMULATED DATA.



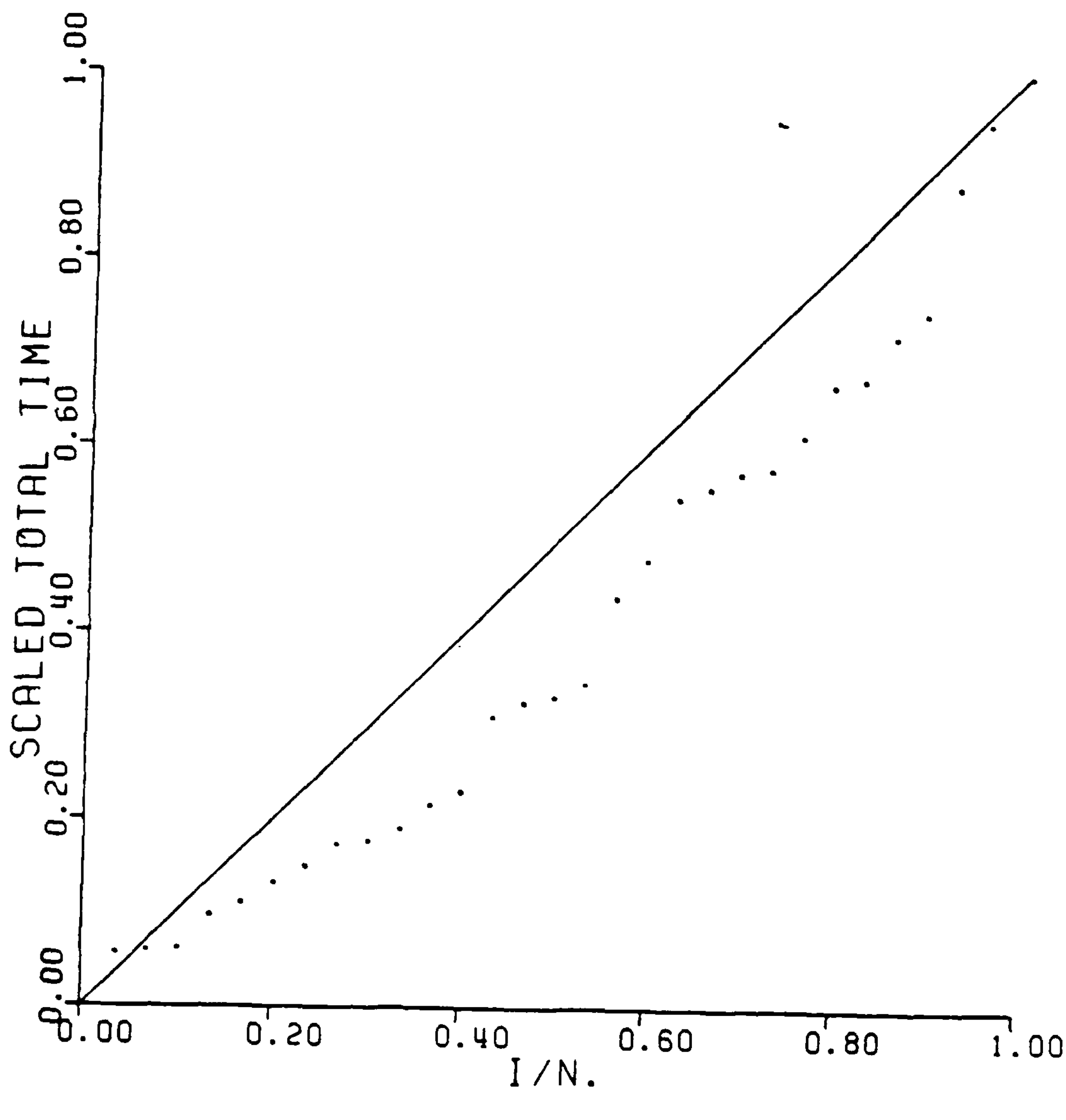


FIGURE 7 B

SCALED TOTAL TIME ON TEST PLOT
FOR SIMULATED DATA.

$\theta_{mle} = .93$	$a_{mle} = 2.98$
$\theta_{mme} = .491$	$a_{mme} = 4.86$
$\theta_{ber} = .720$	$a_{ber} = 7.02$
$\theta_{ls} = .739$	$a_{ls} = 3.58$
$\theta_{wls} = .970$	$a_{wls} = 2.89$

To study the properties of these estimators, a small scale Monte Carlo study was performed. Random samples of size n = 15, 30, 50, 75, or 100 were generated with $\lambda_1 + \lambda_2 = 3$, b = 3, so $\theta = 1$ and a = 2, 3, 5. 1000 samples were generated for each combination of n and a. The bias, standard deviation of the estimates and n, the number of samples where the estimator exists is reported in table 1 for a, table 2 for θ, and in table 3 for an estimator of the system reliability obtained from (3.1) at $t_0 = 9.085$. The true system reliability at t_0 is .8255 when a = 2, .75 when a = 3, and .619 when a = 5. Also reported in each table is the bias and standard deviation of the least square and weighted least square estimators when they are restricted to those samples where the other estimators exist.

From these tables we note that Berger's modified estimator performs very poorly. Also the weighted least squares estimator allows for estimation of parameters in many more samples when n is small. In general the maximum likelihood estimator outperforms the other estimators, however, when the weighted least squares estimator is restricted to those samples where the maximum likelihood estimator exists, this estimator performs much better when n is small. The somewhat better performance of the MLE in terms of bias is deceptive since some of the estimates of a are less than one, which implies that the mean system reiliability is infinite. Also the weighted least squares estimator of system reliability seems to outperform the other estimators of the system reliability in spite of its relatively poor performance as an estimator of θ. Our recommendation is to use the weighted least squares estimator since it more often provides estimators of the relevant parameters and is somewhat easier to compute.

TABLE 1

BIAS AND STANDARD DEVIATION(SD) OF ESTIMATORS OF A

		MAXIMUM LIKELIHOOD			WEIGHTED LEAST SQUARES			LEAST SQUARES			METHOD OF MOMENTS			BERGER'S METHOD		
A	N	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD
2	15	769	4.5	29.	852	4.8	41.	762	6.8	49.	770	9.1	37.	770	14.4	65.
					766	1.3	5.	715	3.5	16.						
2	30	916	2.8	20.	953	4.7	37.	877	6.4	52.	916	6.1	32.	916	10.0	52.
					912	1.1	3.	857	5.4	51.						
2	50	979	5.8	114.	989	1.7	10.	956	4.0	16.	979	8.5	131.	979	15.4	241.
					976	1.0	3.	952	3.5	12.						
2	75	996	0.9	4.	998	1.0	5.	974	2.4	12.	996	2.2	4.	996	4.5	8.
					996	1.0	5.	972	2.4	12.						
2	100	999	0.5	3.	1000	1.7	35.	989	1.5	9.	999	1.7	5.	999	3.7	8.
					999	0.6	2.	989	1.5	9.						
3	15	642	7.3	39.	753	36.4	843.	653	13.1	149.	643	16.8	77.	643	26.0	114.
					636	1.3	9.	573	10.7	159.						
3	30	809	5.7	30.	870	13.8	141.	768	11.9	100.	810	9.9	104.	809	17.7	68.
					804	1.8	6.	731	11.1	102.						
3	50	916	3.6	18.	935	6.9	65.	864	6.4	33.	916	6.6	25.	916	12.5	42.
					912	2.9	29.	851	5.7	32.						
3	75	963	2.5	14.	977	2.8	17.	925	11.6	144.	963	4.7	24.	963	9.6	38.
					958	1.3	5.	923	11.6	144.						
3	100	978	1.7	7.	989	2.1	12.	956	3.7	19.	978	3.0	9.	978	7.2	15.
					978	1.3	5.	952	3.6	19.						
5	15	520	38.7	573.	665	8.2	53.	558	30.3	493.	522	69.8	925.	522	112.9	1505.
					516	-0.7	5.	458	1.8	15.						
5	30	674	20.4	148.	752	9.3	68.	669	13.7	109.	674	31.7	202.	674	56.3	347.
					660	3.2	29.	601	8.4	86.						
5	50	801	7.6	39.	850	9.0	97.	756	13.4	88.	801	11.4	52.	801	23.4	91.
					787	2.0	10.	722	8.6	56.						
5	75	893	12.8	139.	915	8.0	94.	827	6.6	22.	893	15.3	122.	893	32.6	260.
					878	2.9	16.	714	5.8	20.						
5	100	893	9.5	34.	913	19.6	307.	835	11.0	81.	892	13.0	120.	892	27.1	203.
					879	13.7	273.	821	9.5	79.						

TABLE 2

BIAS AND STANDARD DEVIATION(SD) OF ESTIMATORS OF Θ

A	N	MAXIMUM LIKELIHOOD M	BIAS	SD	WEIGHTED LEAST SQUARES M	BIAS	SD	LEAST SQUARES M	BIAS	SD	METHOD OF MOMENTS M	BIAS	SD	BERGER'S METHOD M	BIAS	SD
2	15	769	0.356	1.702	852	-.102	0.742	782	-.192	0.729	770	-.683	0.205	770	-.803	0.122
					766	-.027	0.025	715	-.156	0.701						
2	30	916	0.112	0.919	953	-.135	0.580	877	-.254	0.586	916	-.623	0.192	916	-.798	0.095
					912	-.100	0.567	857	-.239	0.584						
2	50	979	0.016	0.648	989	-.126	0.492	956	-.263	0.514	979	-.575	0.184	979	-.792	0.079
					976	-.115	0.486	952	-.260	0.513						
2	75	996	-.025	0.522	998	-.125	0.432	974	-.247	0.475	996	-.541	0.174	996	-.790	0.065
					996	-.124	0.431	972	-.246	0.474						
2	100	999	-.019	0.437	1000	-.101	0.381	989	-.216	0.423	999	-.508	0.153	999	-.785	0.052
					1000	-.101	0.381	989	-.216	0.423						
3	15	642	0.691	1.900	753	0.210	1.049	653	0.119	1.031	683	-.513	0.348	643	-.705	0.203
					636	0.390	1.040	573	0.245	1.038						
3	30	809	0.175	1.049	870	0.000	0.757	769	-.096	0.745	810	-.469	0.338	809	-.725	0.160
					804	0.074	0.740	731	-.057	0.743						
3	50	916	0.075	0.766	935	-.012	0.618	864	-.112	0.663	916	-.404	0.333	916	-.718	0.135
					912	0.011	0.609	851	-.100	0.660						
3	75	963	0.030	0.624	977	-.027	0.555	925	-.144	0.603	963	-.375	0.322	963	-.717	0.120
					958	-.010	0.546	923	-.142	0.603						
3	100	978	-.028	0.515	989	-.075	0.472	956	-.165	0.511	978	-.345	0.297	978	-.716	0.104
					978	-.055	0.465	952	-.161	0.509						
5	15	522	1.352	3.109	665	0.713	1.609	558	0.578	1.536	522	-.238	0.601	522	-.546	0.348
					516	1.084	1.599	458	0.827	1.561						
5	30	674	0.558	1.609	752	0.366	1.118	669	0.180	1.026	674	-.199	0.576	674	-.584	0.282
					660	0.523	1.104	601	0.286	1.029						
5	50	801	0.256	1.559	850	0.184	0.869	756	0.105	0.863	801	-.193	0.541	801	-.615	0.232
					787	0.267	0.850	722	0.150	0.858						
5	75	893	0.129	0.817	915	0.112	0.728	827	0.014	0.747	893	-.189	0.535	893	-.628	0.210
					878	0.153	0.715	814	0.028	0.744						
5	100	892	0.033	0.683	913	0.020	0.628	835	-.064	0.666	892	-.206	0.494	892	-.644	0.185
					879	0.055	0.615	821	-.050	0.663						

TABLE 3

BIAS AND STANDARD DEVIATION(SD) OF ESTIMATORS OF SYSTEM RELIABILITY AT T=.9085

		MAXIMUM LIKELIHOOD			WEIGHTED LEAST SQUARES			LEAST SQUARES			METHOD OF MOMENTS			BERGER'S METHOD		
A	N	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD	M	BIAS	SD
2	15	769	-.012	.0647	852	-.004	.0586	762	0.002	.0588	770	0.037	.0503	770	0.064	.0463
					766	-.006	.0577	715	0.002	.0586						
2	30	916	-.005	.0473	953	0.002	.0424	877	0.010	.0434	916	0.037	.0357	916	0.069	.0335
					912	0.001	.0426	857	0.009	.0434						
2	50	979	-.001	.0372	989	0.003	.0349	956	0.010	.0359	979	0.035	.0300	979	0.071	.0233
					976	0.003	.0348	952	0.010	.0359						
2	75	996	0.000	.0290	998	0.004	.0274	974	0.010	.0292	996	0.034	.0244	996	0.072	.0238
					996	0.003	.0274	972	0.010	.0293						
2	100	999	0.001	.0243	1000	0.004	.0234	989	0.010	.0248	999	0.034	.0223	999	0.075	.0216
					999	0.004	.0233	989	0.010	.0248						
3	15	642	-.018	.0815	753	-.010	.0767	653	-.006	.0748	643	0.031	.0661	643	0.067	.0616
					636	-.015	.0764	573	-.008	.0751						
3	30	809	-.007	.0577	870	-.003	.0552	769	0.002	.0551	810	0.024	.0490	809	0.062	.0472
					804	-.006	.0544	731	0.001	.0550						
3	50	916	-.003	.0429	935	-.001	.0412	864	0.005	.0431	916	0.022	.0366	916	0.066	.0340
					912	-.002	.0411	851	0.005	.0432						
3	75	963	-.001	.0372	977	0.000	.0356	925	0.007	.0374	963	0.021	.0327	963	0.067	.0313
					958	0.000	.0357	923	0.007	.0375						
3	100	9788	0.001	.0309	989	0.002	.0299	956	0.008	.0311	978	0.019	.0267	978	0.067	.0261
					978	0.002	.0299	952	0.008	.0312						
5	15	520	-.029	.1011	665	-.024	.0967	558	-.020	.0977	522	0.012	.0926	522	0.054	.0921
					516	-.030	.0968	458	-.022	.0968						
5	30	674	-.022	.0691	752	-.020	.0674	669	-.013	.0651	674	0.001	.0613	674	0.045	.0608
					660	-.027	.0665	601	-.016	.0652						
5	50	801	-.006	.0545	850	-.006	.0530	756	-.002	.0532	801	0.010	.0494	801	0.055	.0485
					787	-.008	.0528	722	-.002	.0535						
5	75	893	-.005	.0442	915	-.005	.0436	827	-.002	.0431	893	0.007	.0406	893	0.051	.0380
					878	-.006	.0430	814	-.002	.0432						
5	100	892	-.002	.0375	913	-.001	.0392	835	0.003	.0380	892	0.008	.0353	892	0.052	.0345
					879	-.002	.0390	821	0.003	.0381						

Acknowledgment

This work was supported by the U.S. Air Force Office of Scientific Research under contract AFOSR-82-0307.

Bibliography

Bain, L. J. (1978). Statistical Analysis of Reliability and Life-Testing Models. New York: Marcel Dekker, Inc.

Barlow, R. E. and Campo, R. (1975). Total Time on Test Processes and Applications to Failure Data Analysis. Reliability and Fault Tree Analysis. (Barlow, Fussell, and Singpurwalla, eds.). SIAM, Philadelphia, 451-481.

Basu, A. P. and Klein, J. P. (1982). Some Recent Results in Competing Risk Theory. Survival Analysis (Crowley and Johnson, eds.). The Institute of Mathematical Statistics, 216-227.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familiar Tendency in Chronic Disease Incidence. Biometrica 65, 141-151.

Hui, S. L. and Berger, J. (1983). Empirical Bayes Estimation of Rates. Journal of the American Statistical Association 78, 753-760.

Hutchinson, T. P. (1981). Compound Gamma Bivariate Distributions. Metrika 28, 263-271.

Klein, J. P. and Moeschberger, M. L. (1984). Consequences of Departure from Independence in Exponential Series Systems. Technometrics 26, 277-284.

Lindley D. V. and Singpurwalla, N. D. (1985). Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment. George Washington University Technical Report.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. Journal of the Royal Statistical Society B, 44, 414-422.

Takahasi, K. (1965). Note on the multivariate Burr's Distribution. Annals of the Institute of Statistical Mathematics, Tokyo, 17, 257-260.

Appendix E

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Dependent Competing Risks		5. TYPE OF REPORT & PERIOD COVERED 10/1/84 to 10/31/85
		6. PERFORMING ORG. REPORT NUMBER 714837/763265
7. AUTHOR(s) John P. Klein and Sukhoon Lee		8. CONTRACT OR GRANT NUMBER(s) Grant No. AFOSR 82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Rd., Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 92
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington, D.C. 20332		12. REPORT DATE November 26,1985
		13. NUMBER OF PAGES 2
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington, D.C. 20332		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

On Dependent Competing Risks

by
John P. Klein and Sukhoon Lee
Department of Statistics
The Ohio State University
Columbus, Ohio 43210 USA

1. Introduction

The problem of competing risks arises naturally in a number of contexts, namely the modeling of series systems in reliability, the problem of estimation with censored data and the analysis of physical or biological systems with multiple failure modes. A common, untestable assumption is usually made that the potential failure times for each risk are statistically independent (see Basu and Klein (1982)). Moeschberger and Klein (1984) show that an investigation may be appreciably misled in modeling series systems reliability and in estimating component parameters by incorrectly assuming independent component lifetimes. In this paper we model dependence between components through a common environmental effect on each component. Such a dependence structure has also been suggested by Oakes (1982), Lindley and Singpurwalla (1985), and Hutchinson (1981).

2. The Model

Consider a two-component system. Suppose that under ideal, controlled conditions, as encountered in the testing stage, the times to failure of the two components are X_0 and Y_0. Under these conditions, X_0, Y_0 are independent with marginal survival functions, F_0 and G_0. Now suppose the two components are linked into a system and exposed to the environment. The effect of the environment is to select a random factor, Z, from a distribution, H(z), which changes the marginal survival functions of the two components to $F_0^{\,Z}$, $G_0^{\,Z}$, respectively. A value of Z less than one means that component reliability is improved, while a value greater than one implies a joint degradation. In the sequel we assume that X_0 and Y_0 follow a Weibull distribution with parameters (α_X, λ_X), and (α_Y, λ_Y). That is, $F_0(x) = \exp(-\lambda_X x^{\alpha_X})$. The resulting joint reliability of the two components' lifetimes, (X,Y), in the operating environment is $F(x,y) = E(\exp(-Z(\lambda_X x^{\alpha_X} + \lambda_Y y^{\alpha_Y})))$. F(x,y) is positive quadrant dependent. Also $E(X) = E(X_0)E(Z^{-1/\alpha_X})$; $V(X) = E(X_0^{\,2})E(Z^{-2/\alpha_X}) - (E(X_0)E(Z^{-1/\alpha_X}))^2$; and $Cov(X,Y) = E(X_0)E(Y_0)\,Cov(Z^{-1/\alpha_X}, Z^{-1/\alpha_Y})$. The correlation is always positive and is bounded above by $\Gamma(1 + 1/\alpha)^2/\Gamma(1 + 2/\alpha)$ when $\alpha_x = \alpha_y = \alpha$. Explicit, though lengthy, formula for the moments of (X,Y), system and component reliability, and system components can be obtained when Z is assumed to be either a uniform or gammar random variables. The gamma case leads to bivariate Burr distributions.

3. Estimation

The estimation of model parameters is carried out under the assumption that Z has a gamma distribution with density $h(z) \alpha z^{a-1} \exp(-bz)$. Since the parameters λ_x, λ_y are not identifiable when only data from either a series or parallel system is available, we incorporate sample information obtained independently on each component under test conditions. Maximum likelihood and method of moments estimators are obtained and their properties are studied by Monte-Carlo methods since no closed form maximum likelihood estimates are available.

Acknowledgement

This research was supported by contract AFOSR-82-0307 for the U.S. Air Force Office of Scientific Research.

BIBLIOGRAPHY

1. Basu, A. P. and Klein, J. P. (1982). "Some Recent Results in Competing Risk Theory," in Survival Analysis, Crowley and Johnson, eds., IMS, 216-271.

2. Hutchinson, T. P. (1981). "Compound Gamma Bivariate Distributions," Metrika, 28, 263-271.

3. Lindley, D. V. and Singpurwalla, N. D. (1985). "Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment," Proceedings of the International Conference on Reliability and Quality Control (to appear).

4. Moeschberger, M. L. and Klein, J. P. (1984). "Consequences of Departure from Independence in Exponential Series Systems," Technometrics, 26, 277-284.

5. Oakes, D. (1982). "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society B*, 44, 414-422.

Resume

Un modèle pour les systèmes dépendants dans l'analyse de la fiabilité est examiné. Le modèle suppose que, sous les conditions idéales, les temps de survie des constituants du système ont des distributions Weibull indépendantes. Sous des conditions d'opération un facteur extérieur aléatoire affecte chaque constituant simultanément en multipliant son taux de hasard par une quantité aléatoire. Les propriétés de ce modèle et l'estimation des paramètres du modèle sont considérés, à partir des exemples concrets du laboratoire et de la pratique.

END

1-87

DTIC